Sampling Strategies and Reconstruction Techniques for Magnetic Resonance Imaging

Nicholas D. Blakeley, B.E.(Hons. I), B.Com.

Department of Electrical and Computer Engineering

A thesis presented for the degree of
Doctor of Philosophy

University of Canterbury
Christchurch, New Zealand
February 2003

I am enough of an artist to draw freely upon my imagination. Imagination is more important than knowledge. Knowledge is limited. Imagination encircles the world.

Albert Einstein

And in the end, the love you take is equal to the love you make.

The Beatles

Abstract

In magnetic resonance imaging (MRI), samples of the object's spectrum are measured in the spatial frequency domain (k-space). For a number of reasons there is a desire to reduce the time taken to gather measurements. The approach considered is to sample below the Nyquist density, using prior knowledge of the object's support in the spatial domain to enable full reconstruction.

The two issues considered are where to position the samples (sampling strategies) and how to form an image (reconstruction techniques). Particular attention is given to a special case of irregular sampling, referred to as Cartesian sampling, in which the samples are located on a Cartesian grid but only constitute a subset of the full grid. A further special case is considered where the sampling scheme repeats periodically, referred to as periodic Cartesian sampling. These types of sampling schemes are applicable to 3-D Cartesian MRI, MRSI, and other modalities that measure a single point in 2-D k-space per echo. The case of general irregular sampling is also considered, which is applicable to spiral sampling, for example.

A body of theory concerning Cartesian sampling is developed that has practical implications for how to approach the problem and provides intuition about its nature. It is demonstrated that periodic Cartesian sampling effectively decomposes the problem into a number of much smaller subproblems, which leads to the development of a reconstruction algorithm that exploits these computational advantages. An additional algorithm is developed to predict the regions that could be reconstructed from a particular sampling scheme and support; it can be used to evaluate candidate sampling schemes before measurements are obtained. A number of practical issues are also discussed using illustrative examples. Sample selection algorithms for both Cartesian and periodic Cartesian sampling are developed using heuristic metrics that are fast to compute. The result is a significant reduction in selection time at the expense of a slightly worse conditioned system. The reconstruction problem for a general irregular sampling scheme is also analysed and a reconstruction algorithm developed that trades off computation time for better image quality.

Acknowledgements

As a prospective postgraduate student, the pearls of wisdom I gained from current students were that a good relationship with your supervisor is essential. I have been lucky enough to have two great relationships: I am enormously grateful to Phil Bones and Rick Millane for their time, effort, guidance, and friendship. I am also grateful to Peter Renaud for his enthusiasm and ideas (and for allowing engineers into the mathematics department).

In the course of my research I enjoyed a short stay at the Robarts Research Institute at the University of Western Ontario in London, Ontario, Canada. I express my thanks to the Institute for allowing me to visit and to obtain some MRI data (at the University Hospital). I am especially grateful to Atsushi Takahashi and Yi-Fen Yen, who taught me about MRI (I learnt a *lot* in those two weeks!), showed me the MRI scanners, and made my stay in London very enjoyable both at and away from work. I also thank Terry Peters for discussions and ideas, and him and Jackie for being so friendly and welcoming. In addition, my thanks go to Jeremy Mozzon for showing me around the university and teaching me about strange Canadian sports.

I gratefully acknowledge the financial support provided by the Foundation for Research, Science & Technology. In particular, the allowance for conference travel has been an invaluable part of the scholarship.

Finally, I would like to thank my friends and family, especially Ingrid (where would I be without you), for ensuring my sanity and basically making life an enjoyable experience.

Contents

Preface

The subject of this thesis is image reconstruction of a spatial domain object with limited support from data sampled in the spatial frequency domain. This problem arises in many situations, but magnetic resonance imaging (MRI) is the particular application considered and the source of many examples. The two main aspects of the problem that are investigated are sampling strategies and reconstruction techniques.

In Chapter 1, the topic is introduced and motivation provided for the proposed sampling approach, namely sampling below the Nyquist density to reduce acquisition time in MRI. The consequences of such an approach to the MRI process are outlined and the main objectives for the research are discussed.

Some of the relevant mathematical foundations for the methods in the thesis are presented in Chapter 2. In particular, the notation and adopted indexing convention are defined and the MRI reconstruction problem is discussed in relation to the general field of inverse problems.

Because of the emphasis on MRI as an application, an overview of MRI theory and practice is given in Chapter 3. The simulated phantom and *in vivo* data used for examples in later chapters are also discussed and the problem of establishing the region of support is addressed.

Original contributions are contained in Chapters 4 to 7. The two main types of sampling considered in the thesis are irregular sampling (IS) and Cartesian sampling (CS). The term Cartesian sampling is used in this thesis to describe the special case of irregular sampling in which the samples lie on a regular Cartesian grid and a subset of the full Nyquist set is obtained. A further special case of Cartesian sampling is considered where the sampling set repeats periodically: periodic Cartesian sampling (PCS).

Although a special case, Cartesian sampling is interesting for a number of reasons: it is relevant to applications in MRI, it represents the case in other sampling procedures where samples from a complete grid are missing or corrupted, and it allows use of the discrete

Fourier transform. The majority of the research concerns Cartesian sampling, which is the specific topic of Chapter 4. A mathematical framework is established and a body of theory developed that highlights some important properties of Cartesian sampling.

The case of periodic Cartesian sampling is studied in Chapter 5 and it is shown to simplify the problem by allowing decomposition into smaller subproblems; the relationship between the subproblems and the overall problem is demonstrated mathematically. The computational benefits are exploited in the development of both a reconstruction algorithm and a method for predicting the success of a reconstruction.

In Chapter 6, the task of sample selection is considered for both CS and PCS. Universal patterns are used as a basis for approaching selection in the case of PCS. Emphasis is placed on speed of selection and this is achieved through the development of heuristic metrics that do not rely on computationally intensive linear algebraic techniques. These metrics are incorporated into selection algorithms for both types of sampling.

The general case of irregular sampling is investigated in Chapter 7, culminating in the development of a reconstruction algorithm that trades off speed for better image quality. The method can also be applied to Cartesian sampling, in which case the system matrix has specific structure that can be exploited in the method of solution.

Finally, conclusions and suggestions for future work are discussed in Chapter 8.

The novel contributions contained in the thesis are summarised in the introduction to each of Chapters 4 to 7, and are also listed below:

- The trace metric appears elsewhere in the literature, but the derivation presented here is more rigorous and also includes an analysis of fair comparisons.
- The relationship between 1-D and 2-D sequences appears in the literature, but it is proved here for 1-D and M-D sequences and is described explicitly.
- Results on cyclic shifts and mirror images appear elsewhere in the literature, but the equivalence is proved here, a related new result is obtained, and the number of 'unique' sampling sets is calculated precisely.
- Results are presented on the singular values of the problem, including an application of a previous result, a relatively straightforward property, and an interesting result on complementary problem sizes.
- A conjecture concerning the optimality of a maximally spread sampling pattern for a contiguous support is presented and proved for certain conditions .

- Results on universal patterns appear elsewhere in the literature, but an interesting new result on equivalent universal patterns is presented here.
- A consequence of the maximally spread sampling pattern conjecture is investigated and discussed using an example.
- It is argued by way of example that larger problems are generally worse conditioned than smaller problems for a similar ratio of support size to sequence length.
- A reconstruction technique for periodic Cartesian sampling is presented which is essentially the same as another that appears in the literature, although it is developed for M-D signals.
- A method for predicting the success of reconstruction for periodic Cartesian sampling is developed.
- An interesting result is presented concerning the distribution of singular values of the overall problem among the subproblems.
- The utility of periodic Cartesian sampling is discussed in terms of sampling density and conditioning and it is argued that periodic Cartesian sampling is not disadvantageous compared to general Cartesian sampling.
- It is argued that singular systems are caused by distances between sample locations that are factors of the size of the discrete sequences, and two heuristically-based metrics based on this contention are developed.
- Alternative measures of conditioning are proposed that are fast to compute, leading to the development of two heuristically-based metrics.
- A fast, sequential-selection algorithm is presented for Cartesian sampling.
- A fast, sequential-selection algorithm is presented for periodic Cartesian sampling.
- A reconstruction technique for irregular sampling is presented, although it is of limited novelty: a number of existing results are brought together and applied to the MRI problem. The technique is very similar to another that appears in the literature.

The journal and conference papers that have resulted from the research described in the thesis are listed below in chronological order. The first two papers are mainly concerned with the PCSR algorithm developed in Chapter 5; the third paper covers the ISR algorithm described in Chapter 7; and the final three papers consider the PCSS algorithm utilising the rapid selection of universal patterns, as discussed in Chapter 6.

N. D. Blakeley, N. Alwesh, T. J. Connolly, and P. J. Bones. "Recovery of limited extent images from regularly undersampled spectra," in *Proceedings of the Image & Vision Computing New Zealand conference (IVCNZ)*, Hamilton, New Zealand, November 2000, pp. 44–49.

Philip J. Bones, Nawar Alwesh, T. John Connolly, and Nicholas D. Blakeley. "Recovery of limited-extent images aliased because of spectral undersampling," *Journal of the Optical Society of America A*, vol. 18, no. 9, pp. 2079–2088, September 2001.

P. J. Bones, N. D. Blakeley, and R. P. Millane. "Image recovery from irregularly located spectral samples," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, Thessaloniki, Greece, October 2001, vol. 1, pp. 217–220.

N. D. Blakeley and P. J. Bones. "Full recovery of limited extent images from regularly undersampled spectra," in *Proceedings of the Image & Vision Computing New Zealand conference (IVCNZ)*, Dunedin, New Zealand, November 2001, pp. 45–50.

N. D. Blakeley, R. P. Millane, P. Renaud, and P. J. Bones. "Efficient selection of k-space samples for recovering limited-extent images," in *Proceedings of the International Society of Magnetic Resonance in Medicine (ISMRM)*, Honolulu, Hawai'i, USA, May 2002, p. 2434.

Nicholas D. Blakeley, P. J. Bones, R. P. Millane, and Peter Renaud. "Efficient frequency-domain sample selection for recovering limited-support images," *Journal of the Optical Society of America A*, vol. 20, no. 1, pp. 67–77, January 2003.

Abbreviations

2DFT	2-D Fourier transform (MR technique)
A/D	analog-to-digital
BCCB	block circulant with circulant blocks
BOLD	blood oxygen level dependent (MR technique)
BTTB	block Toeplitz with Toeplitz blocks
CS	Cartesian sampling
CSS	Cartesian sample selection (algorithm)
CSS1	Cartesian sample selection 1-D (algorithm)
CT	computed tomography
DFT	discrete Fourier transform
EPI	echo planar imaging (MR technique)
FFT	fast Fourier transform
FID	free induction decay
FISP	fast imaging with steady-state precession (MR technique)
FLASH	fast low angle shot (MR technique)
fMRI	functional magnetic resonance imaging

FOV	field of view
FSE	fast spin echo (MR technique)
GE	gradient echo (MR technique)
G-P	Gerchberg-Papoulis
GRASS	gradient refocussed acquisition in the steady state (MR technique)
IDFT	inverse discrete Fourier transform
IFFT	inverse fast Fourier transform
IS	irregular sampling
ISR	irregular sampling reconstruction (algorithm)
LUT	lookup table
MR	magnetic resonance
MRI	magnetic resonance imaging
MRS	magnetic resonance spectroscopy
MRSI	magnetic resonance spectroscopic imaging
NMR	nuclear magnetic resonance
PCS	periodic Cartesian sampling
PCSP	periodic Cartesian sampling prediction (algorithm)
PCSR	periodic Cartesian sampling reconstruction (algorithm)
PCSS	periodic Cartesian sample selection (algorithm)
PET	positron emission tomography
POCS	projection onto convex sets
RF	radiofrequency
RMS	root mean square
ROI	region of interest
SBS	sequential backward selection
SE	spin echo (MR technique)
SENSE	sensitivity encoding (MR technique)
SFS	sequential forward selection
SMASH	simultaneous acquisition of spatial harmonics (MR technique)
SNR	signal-to-noise ratio
SSFP	steady state free precession (MR technique)
SVD	singular value decomposition
UNFOLD	unaliasing by Fourier-encoding the overlaps using the temporal dimension (MR technique)
UPS1	universal pattern selection 1-D (algorithm)
WKS	Whittaker-Kotel'nikov-Shannon

Chapter 1

Introduction

Modern medical imaging techniques such as computed tomography (CT) and magnetic resonance imaging (MRI) reveal remarkable detail of the internal structure of human beings. They have enhanced doctors' abilities to diagnose illness and provided insight into the mechanisms of the living brain. What remains unseen to the viewer, however, is the elegant theory and practical mathematical procedures that make it possible to turn the measured data into an image. In x-ray imaging or conventional photography, the image is ready to be viewed as soon as the measurements are collected—the object is measured directly by the imaging process. However, in CT and MRI this is not the case and the data must be manipulated computationally to form an image.

This type of problem is found throughout the field of imaging in applications such as radio astronomy or image deblurring, and it lies within the more general field of inverse problems: determining unknown causes from known effects. In the case of MRI, the unknown causes are the properties of tissues in the body and the known effects are the radiofrequency signals measured. While it is relatively straightforward to predict the measurements if the tissue properties are known, solving the problem in the reverse direction is much more difficult.

Note that certain terminology is introduced in the remainder of this chapter with only brief description. These terms are more fully explained in Chapters 2 and 3 with references to appropriate supporting literature.

1.1 Limitations in magnetic resonance imaging

In many imaging modalities, it is necessary or convenient to measure the Fourier transform of an object rather than the object itself, and this is the case in MRI, CT and radio astronomy.

However, in magnetic resonance (MR) great freedom exists in choosing the locations of the measurements (or samples) in the spatial frequency (or Fourier) domain, known in the MR community as k-space. The most common sampling procedure is to acquire samples on a regular Cartesian grid, with sample spacings dictated by the Nyquist criterion. A sampling scheme such as this in which all samples within a rectangular block (or in M dimensions, an M-dimensional cuboid block) surrounding the origin are measured is referred to as a full Nyquist set. The size of the rectangular block determines spatial resolution and the prevalence of Gibbs ringing, while the fact that the complete block is measured ensures aliasing does not occur. MR measurements are obtained from echoes following excitation by a radiofrequency pulse; samples for a complete line of k-space can be measured per echo. The enormous benefit of this scheme is that it allows use of the fast Fourier transform (FFT) to reconstruct an image. While a full Nyquist set is often convenient to obtain, there are a number of reasons why one may wish to depart from this sampling scheme. In particular, it can suffer from long acquisition time—a perennial limitation in MRI. To understand its impact on the imaging process, three of the key parameters in MRI are described here: acquisition time, signal-to-noise ratio (SNR), and resolution.

Acquisition time: The duration of an MR 'scan' depends on how much data is to be collected: the more measurements, the longer the scan. A scan consists of a series of echoes, each of which may contain information for a number of samples in k-space. Each echo takes of the order of a few milliseconds to collect. As might be expected, a better image can be obtained from more data, but there is a continual push by clinicians to reduce the scan time for a number of reasons, discussed below.

SNR: One way of measuring the quality of an image is the signal-to-noise ratio. This can be affected by a number of imaging parameters but in general a shorter scan time means a decrease in SNR. How crucial SNR is depends on the application.

Resolution: The smallest feature that can be distinguished in an image is governed by the spatial resolution, which is a function of the extent of k-space covered by the measurements. Of course, taking more measurements in order to cover a greater extent of k-space incurs an increase in acquisition time.

The benefits derived from a reduction in scan time can be grouped into three categories: throughput, stationarity constraints, and feasibility. Firstly and most obviously, a shorter scan time means that more scans can be completed per hour and the expensive scanner itself is therefore used more efficiently. Although a two-dimensional (2-D) scan of a full Nyquist set can be completed in the order of ten seconds, faster scanning could be used to provide a more complete picture of the same patient (with more image 'slices') or to allow a greater number of patients access to what is still a scarce resource (although use of MRI will hopefully become more widespread as the cost of machines decreases).

Three-dimensional (3-D) MRI suffers from very long scan times (a 3-D scan of the author took twelve minutes), during which time the patient must remain completely still. The long scan time increases the likelihood of movement artefacts, and because the measurements are gathered in k-space, a small movement during one of the measurements can adversely affect the entire image. Patient welfare is also a priority and many patients have injuries or disabilities (such as arthritis) that may make it quite uncomfortable for them to remain motionless in the scanner for an extended period of time. This is relevant for both a single long scan or a number of shorter scans. Reducing scan times therefore has the potential to alleviate movement artefacts and patient discomfort.

Long scan times are also a problem in magnetic resonance spectroscopic imaging (MRSI). This imaging technique exploits the chemical shift phenomenon to provide cellular metabolic information, but it is inherently slower than conventional MRI since a single echo must be used for every location in k-space. A Cartesian grid is normally used, but to avoid very long scan times a relatively small portion of k-space is measured, resulting in low spatial resolution and artefacts (Gibbs ringing). As an alternative to simply reducing scan time, in this case it may be desirable to use any scan time saved to gather more measurements and increase resolution (and reduce ringing).

Finally, while conventional structural MRI is concerned with the gross anatomical structures of the body, other MR applications involve events that occur in real-time. For example, cinematic (or cine) imaging of the cardiac cycle allows heart conditions to be evaluated, while functional imaging of the neuronal activity in the brain can be used to study mental disorders. To make these methods feasible, scan times must be much shorter than a second.

1.2 Solutions

The desire to reduce acquisition time has led to the advent of fast-scan techniques, where the same amount of data is collected in a shorter time. Some of the methods developed result in sample locations that do not necessarily lie on a Cartesian grid; for example, spiral sampling (named after its trajectory in k-space) is a popular technique in which an entire scan can be completed in a single echo. In cases such as this, however, specialised reconstruction techniques are required to form an image.

A second approach is reduced-scan, which involves obtaining less data than is ordinarily required and exploiting prior knowledge about the object to produce an image. For example, the spectrum of a purely real object (as found in MRI) exhibits Hermitian symmetry, which means that only half of the spectrum needs to be measured. The penalty for reduced-scan techniques is generally a reduction in SNR; in applications where this is not crucial, these techniques provide a tradeoff between acquisition time and SNR.

The objectives of the research presented in this thesis were to investigate sampling strategies that result in a reduction in acquisition time and to develop associated algorithms for image reconstruction. These aims may be summarised by the following questions:

1. How few samples can be measured while still enabling a good image to be formed?
2. Where should the samples be placed?
3. What reconstruction algorithm should be used?

The emphasis on taking fewer samples to reduce acquisition time places the majority of the methods investigated in this thesis into the reduced-scan category, although reconstruction for fast-scan methods such as spiral sampling is also considered. The sampling schemes investigated can be divided into two main types: Cartesian and irregular sampling. Irregular sampling encompasses schemes such as spiral sampling in which the samples are not regularly spaced with respect to a Cartesian coordinate system. The term Cartesian sampling is used in the thesis to describe the special case of irregular sampling in which the samples are located on a Cartesian grid but only constitute a subset of the full Nyquist set.

There are a number of forms of prior knowledge that can be used in MRI. The primary quantity measured in the MR imaging process is the number of protons per unit volume (proton density), which clearly must be a real and non-negative quantity. The former results in the spatial frequency domain property of Hermitian symmetry, as mentioned above, while the latter provides the spatial domain property of positivity. A statistical prior is found in the histogram of differences in intensity between adjacent image pixels, which is modelled by a Lorentzian distribution. However, the main type of prior information used in this thesis is limited support: the object is only *supported* (non-zero) on a particular region or regions in the spatial domain. To be sure, sampling theory requires that the object have finite extent, but a full Nyquist set in the spatial frequency domain implies a corresponding rectangular field of view (FOV) in the spatial domain within which the object may lie without its reconstructed image incurring aliasing. In some cases there are regions *within* the FOV that are known *a priori* to be identically zero; for example, a cross-sectional image of a person's head can only be non-zero in the region lying within the head. In imaging of blood flow through arteries, there may be several distinct regions within which the object is non-zero (or alternatively the object differs from some other known fixed value). With such additional information, it is possible to omit some samples from the full Nyquist set while still allowing the formation of an image. This approach is applicable to 3-D Cartesian MRI, MRSI, and other modalities that measure a single point in 2-D k-space per echo.

To summarise the alternative imaging procedure being investigated here, consider the steps that comprise forming an MR image conventionally:

1. Take measurements at sampling locations covering a full Nyquist grid.
2. Form an image using the fast Fourier transform.

To implement the proposed reductions in acquisition time using prior knowledge of the support, the following structure would be necessary:

1. Find the support.
2. Determine an appropriate sampling pattern.
3. Take measurements at the selected sampling locations.
4. Form an image using an appropriate reconstruction technique.

Note that the support region generally varies from scan to scan (because of different patient geometries, different anatomical sections being imaged, etc.) so the best sampling pattern needs to be calculated for each scan.

The motivation for this more complicated structure is a reduction in acquisition time and this is realised by taking fewer measurements compared to a conventional acquisition (step three). However, the price to be paid is additional pre- and post-processing. In an MRI examination, it is usual to obtain several preliminary images (a fast procedure) and standard image processing techniques could then be used to establish the support (step one). Selecting the sampling pattern itself (step two) is not as easy, and speed is clearly a priority since the samples cannot be measured until they are selected. Reconstruction of an image from the resulting set of samples (step four) also requires a specialised procedure, although speed is not quite so critical.

It should be noted that under this structure there is an unambiguous increase in processing time; it could even be the case that the overall imaging time (i.e. the combination of processing and acquisition) actually increases. Notwithstanding, the benefits realised from reduced acquisition time may make this a price worth paying, especially considering computer processing is so cheap compared to the cost of an MR scanner.

Although discussed here in relation to MR imaging, the problem described above is in fact more general, namely reconstruction of a spatial domain object with limited support sampled in the spatial frequency domain. An inversion of the domains has no effect on the problem, so this is equivalent to reconstruction of a bandlimited object sampled in the spatial domain. It is partly this generality that makes the problem an interesting and important one to study.

Chapter 2

Mathematical preliminaries

To provide a foundation for the work presented later in the thesis, this chapter covers some of the relevant mathematical theory. For easy reference, the notation used is summarised in Table 2.1.

2.1 Fourier transforms and discrete sequences

The **Fourier transform** lies at the heart of much electrical engineering, and especially signal processing [Bra65]. An M-dimensional (M-D) signal $f(\mathbf{x})$, where $\mathbf{x} = [x_0, x_1, \ldots, x_{M-1}]^T \in \mathbb{R}^M$, has Fourier transform

$$F(\mathbf{u}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(\mathbf{x}) \exp(-i2\pi \mathbf{u}^T \mathbf{x}) d\mathbf{x}, \tag{2.1}$$

where $\mathbf{u} = [u_0, u_1, \ldots, u_{M-1}]^T \in \mathbb{R}^M$ represents a position in frequency space, and $d\mathbf{x} = dx_0 dx_1 \cdots dx_{M-1}$. The inverse Fourier transform is given by

$$f(\mathbf{x}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} F(\mathbf{u}) \exp(i2\pi \mathbf{u}^T \mathbf{x}) d\mathbf{u}, \tag{2.2}$$

where $d\mathbf{u} = du_0 du_1 \cdots du_{M-1}$.

Once continuous signals are sampled, such as for use on a digital computer, the **discrete Fourier transform** (DFT) must be used [Bri88]. It may be computed via the **fast Fourier transform** (FFT) [CT65, Bri88], which provides a computational advantage that continues to make the FFT one of the most important numerical algorithms today. The DFT operates on finite-length discrete sequences; therefore, consider an M-D discrete sequence $f[\mathbf{n}]$ of size $\mathbf{N} = [N_0, N_1, \ldots, N_{M-1}]^T \in \mathbb{Z}^M$, where $\mathbf{n} = [n_0, n_1, \ldots, n_{M-1}]^T \in \mathbb{Z}^M$. Periodicity is assumed outside this size; that is, $f[\mathbf{n}] = f[\mathbf{n} + \mathbf{m} \circ \mathbf{N}]$ for any $\mathbf{m} \in \mathbb{Z}^M$. The DFT of $f[\mathbf{n}]$ is

Table 2.1 Summary of notation used in the thesis (continued on next page)

symbol	meaning
i	$\sqrt{-1}$
$\mathbb{Z}$	the set of integers
$\mathbb{R}$	the set of real numbers
$\mathbb{C}$	the set of complex numbers
$\|\cdot\|$	absolute value; magnitude of a complex number; size of a set
$\bar{x}$	complex conjugate; set complement
$x \backslash y$	set difference
$\lfloor \cdot \rfloor$	floor of a real number (i.e. round towards $-\infty$)
A	matrix
$(A)_{ij}$	*(i,j)*th element of A
$\mathbf{x}$	(column) vector
$(\mathbf{x})_i$ or x_i	ith element of $\mathbf{x}$
I	identity matrix
A^T	transpose of A
A^H	Hermitian transpose of A
A^{-1}	inverse of A
A^+	Moore-Penrose inverse (also called the pseudo-inverse) of A
$A[\alpha\|\beta]$	submatrix of A with rows and columns indexed by sets α and β
$A[\alpha\|\cdot]$	submatrix of A with rows indexed by the set α and all columns
$\mathbf{x}[\alpha]$	subvector of $\mathbf{x}$ with rows indexed by the set α
$\|\cdot\|$	2-norm (or spectral norm) of a matrix or vector
$\|\cdot\|_F$	Frobenius norm of a matrix
$\mathcal{N}(A)$	nullspace of A
$\mathrm{trace}(A)$	trace of A
$\det(A)$	determinant of A
$\mathrm{rank}(A)$	rank of A
$\mathrm{cond}(A)$	2-norm condition number of A
$\mathrm{tmetric}(A)$	trace metric of A (see Section 4.4.1)
$\circ$	Hadamard or Schur product (element-by-element multiplication)
$\circ^{-1}$	element-by-element division
$\otimes$	Kronecker product
$\odot$	convolution operator
$\circledast$	correlation operator
$P(\cdot)$	probability operator
$E\{\cdot\}$	expected value operator
$f(\mathbf{x})$, $F(\mathbf{u})$	continuous signal: spatial domain, frequency domain
$f[\mathbf{n}]$, $F[\mathbf{k}]$	discrete sequence: spatial domain, frequency domain

Table 2.1 Summary of notation used in the thesis (continued from previous page)

symbol	meaning
$\mathcal{S}(\mathbf{n}, \mathbf{N})$	stacking index operator (see Section 2.3.1)
$\mathcal{S}^{-1}(l, \mathbf{N})$	inverse stacking index operator (see Section 2.3.1)
W_N	unitary Fourier matrix for one-dimensional sequences of size $N \in \mathbb{Z}$
$W_{\mathbf{N}}$	unitary Fourier matrix for M-dimensional sequences of size $\mathbf{N} \in \mathbb{Z}^M$
$\mathrm{circ}(\mathbf{c})$	circulant matrix with first row given by $\mathbf{c}$
$\mathrm{diag}(\mathbf{d})$	diagonal matrix with elements given by $\mathbf{d}$
$a \bmod m$	modulo operator
$\mathbf{a} \bmod \mathbf{m}$	element-by-element modulo operator ($\mathbf{a}$ and $\mathbf{m}$ are the same size)
$a \mid b$, $a \nmid b$	a divides b, a does not divide b
$(a_1, a_2, \ldots, a_n)$	greatest common divisor of $a_1, a_2, \ldots, a_n$
$a \equiv b \pmod m$	a is congruent to b modulo M

defined as

$$F[\mathbf{k}] = \left(\prod_{i=0}^{M-1} N_i \right)^{-\frac{1}{2}} \sum_{n_0=0}^{N_0-1} \cdots \sum_{n_{M-1}=0}^{N_{M-1}-1} f[\mathbf{n}] \exp\left[-i2\pi \mathbf{k}^T (\mathbf{n} \circ^{-1} \mathbf{N}) \right], \tag{2.3}$$

where $\mathbf{k} = [k_0, k_1, \ldots, k_{M-1}]^T \in \mathbb{Z}^M$ is defined as for $\mathbf{n}$ but represents a position in discrete frequency space. The inverse DFT (IDFT) is then

$$f[\mathbf{n}] = \left(\prod_{i=0}^{M-1} N_i \right)^{-\frac{1}{2}} \sum_{k_0=0}^{N_0-1} \cdots \sum_{k_{M-1}=0}^{N_{M-1}-1} F[\mathbf{k}] \exp\left[i2\pi \mathbf{k}^T (\mathbf{n} \circ^{-1} \mathbf{N}) \right]. \tag{2.4}$$

The circular convolution of two discrete sequences $f[\mathbf{n}]$ and $h[\mathbf{n}]$ is given by

$$f[\mathbf{n}] \odot h[\mathbf{n}] = \sum_{m_0=0}^{N_0-1} \cdots \sum_{m_{M-1}=0}^{N_{M-1}-1} f[\mathbf{m}] h[\mathbf{n} - \mathbf{m}]. \tag{2.5}$$

With the above definition of the DFT, the convolution rule includes a factor of $\sqrt{N}$: if the DFTs of $g[\mathbf{n}]$ and $h[\mathbf{n}]$ are $G[\mathbf{k}]$ and $H[\mathbf{k}]$, respectively, then

$$\begin{aligned} G[\mathbf{k}] &= F[\mathbf{k}] H[\mathbf{k}] \\ \Leftrightarrow g[\mathbf{n}] &= \frac{1}{\sqrt{N}} f[\mathbf{n}] \odot h[\mathbf{n}]. \end{aligned} \tag{2.6}$$

The process of converting a continuous signal into a discrete sequence is governed by sampling theory.

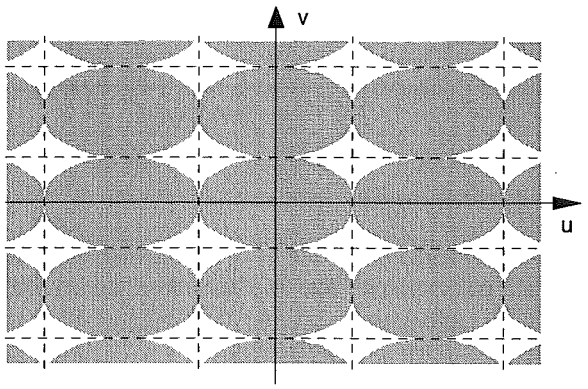


Figure 2.1 The spectral support of a 2-D signal sampled at the Nyquist density.

2.2 Sampling theory

The field of information theory was born with Shannon's 1949 seminal paper [Sha49], which formalised the sampling theorem—"a fact which is common knowledge in the communication art." Refs. [Jer77, Uns00] are review articles of what is now known as the Whittaker-Kotel'nikov-Shannon (WKS) sampling theorem, stated originally by Shannon as follows:

Theorem 2.1 (Shannon [Sha49]). *If a function $f(t)$ contains no frequencies higher than W cps, it is completely determined by giving its ordinates at a series of points spaced $1/2W$ seconds apart.*

In other words, for a given bandwidth, W, a set of uniformly-spaced samples taken at a rate f_s completely determines $f(t)$ provided that $f_s \geq 2W$; the minimum bound is called the Nyquist rate, or density.

For one-dimensional (1-D) low-pass signals, the Nyquist density is equal to the extent of the support of the signal's spectrum. Sampling in the time domain causes periodicity in the frequency domain, and in 1-D, sampling at the Nyquist density produces maximal packing—there are no gaps between spectral replications. However, this is not generally true for band-pass or multi-band 1-D signals, or in higher dimensions. For example, Fig. 2.1 shows the replicated spectral supports of a 2-D signal sampled at the Nyquist density: the roughly elliptical shape of the support has resulted in gaps between the replications. When such gaps exist, the samples can be shown to be linearly dependent, meaning that more than enough samples exist to completely determine the underlying signal [Mar86, CM90].

The fundamental lower limit to sampling density is called the minimum density and was shown by Landau to be the Lebesgue measure of the support [Lan67]. (The Lebesgue measure is an extension to the classical notions of length and area to more complicated

sets; it is the standard way of assigning a volume to subsets of Euclidean space.) This bound, which applies to both uniform and nonuniform sampling, is equal to the Nyquist density in the special case of a 1-D low-pass signal.

In his 1956 paper on nonuniform (or irregular) sampling of a band-limited signal, Yen investigated some particular types of irregular sampling and their reconstruction formulas [Yen56]. In particular, his least-squares optimal bandlimited interpolation algorithm has also been derived using different approaches (e.g. Refs. [CA87, CM00]).

Theorem 2.2 (Yen [Yen56]). *If the sample values at a finite set of arbitrarily distributed [distinct] sample points $t = \tau_p$, $p = 1, 2, \ldots, N$ are given, a signal $f(t)$ with no frequency component above W cps is defined uniquely under the condition that the "energy" of the signal $\int_{-\infty}^{\infty} f(t)^2\,dt$ is a minimum. Moreover, the reconstruction of the signal is*

$$f(t) = \sum_{p=1}^{N} f(\tau_p)\Psi_p(t), \tag{2.7}$$

where

$$\Psi_p(t) = \sum_{q=1}^{N} a_{qp} \frac{\sin 2\pi W(t-\tau_q)}{2\pi W(t-\tau_q)}. \tag{2.8}$$

The coefficients a_{qp} are the coefficients of the inverse matrix of a matrix whose elements are

$$\frac{\sin 2\pi W(\tau_p - \tau_q)}{2\pi W(\tau_p - \tau_q)}, \quad p, q = 1, 2, \ldots, N. \tag{2.9}$$

With a discrete set of samples that results from any sampling procedure, the mathematics of linear algebra may be used for manipulation and analysis.

2.3 Linear Algebra

2.3.1 Indexing convention

Throughout the thesis, indices for matrices and vectors begin at zero. In dealing with M-D discrete sequences, the convention used for stacking sequences into vectors is the same as that used in MATLAB®: an M-D discrete sequence, $f[\mathbf{n}]$, of size $\mathbf{N}$ (as defined in Section 2.1) is stacked along the coordinate governed by N_0, then N_1, and so on. The lth element of $\mathbf{f}$ is $f[\mathbf{n}]$, i.e. $(\mathbf{f})_l = f[\mathbf{n}]$, where the relationship between l and $\mathbf{n}$ is given by

$$l = n_0 + n_1 N_0 + n_2 N_0 N_1 + \cdots + n_{M-1} N_0 N_1 \cdots N_{M-2}, \tag{2.10}$$

or equivalently,

$$\mathbf{n} = \begin{pmatrix} n_0 \\ n_1 \\ n_2 \\ \vdots \\ n_{M-1} \end{pmatrix} = \begin{pmatrix} l \bmod N_0 \\ \lfloor l/N_0 \rfloor \bmod N_1 \\ \lfloor l/(N_0 N_1) \rfloor \bmod N_2 \\ \vdots \\ \lfloor l/(N_0 N_1 \cdots N_{M-2}) \rfloor \bmod N_{M-1} \end{pmatrix}. \tag{2.11}$$

For convenience, define the stacking index operator, $\mathcal{S}(\mathbf{n}, \mathbf{N})$, and the inverse stacking index operator, $\mathcal{S}^{-1}(l, \mathbf{N})$, which apply Eqs. (2.10) and (2.11), respectively. To aid readability, it may be useful to interpret $\mathcal{S}^{-1}(l, \mathbf{N})$ simply as l, which is equivalent to the 1-D case; for example, this converts the imposing expression $h[\mathcal{S}^{-1}(m, \mathbf{N}) - \mathcal{S}^{-1}(n, \mathbf{N})]$ into the more familiar $h[m - n]$.

2.3.2 Singular value decomposition and the Moore-Penrose inverse

For any $A \in \mathbb{C}^{m \times n}$ with rank r, there exist unitary matrices $U \in \mathbb{C}^{m \times m}$ and $V \in \mathbb{C}^{n \times n}$ such that

$$A = U \Sigma V^H, \tag{2.12}$$

called the **singular value decomposition** (SVD) of A [HJ85, Str88, Bar90]. Σ is a real $m \times n$ matrix with all entries zero except the diagonal elements, which are called the *singular values* of A and satisfy

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0, \quad \sigma_{r+1} = \cdots = \sigma_{\min(m,n)} = 0. \tag{2.13}$$

The r non-zero singular values are the square roots of the non-zero eigenvalues of both AA^H and $A^H A$. (Note that AA^H and $A^H A$ are Hermitian and positive semi-definite, so their eigenvalues are non-negative.) The columns of U and V are the eigenvectors of AA^H and $A^H A$, respectively.

For a rectangular linear system, $A\mathbf{x} = \mathbf{b}$, the least-squares solution, $\bar{\mathbf{x}}$, is the vector that minimises $\|A\bar{\mathbf{x}} - \mathbf{b}\|^2$. If the columns of A are linearly independent (i.e. it has full column rank), then $A^H A$ is non-singular and the solution is given by

$$\bar{\mathbf{x}} = (A^H A)^{-1} A^H \mathbf{b}. \tag{2.14}$$

If the columns of A are linearly dependent, then $\bar{\mathbf{x}}$ is not unique and a family of solutions exist that meet the least-squares criterion. To establish uniqueness, the solution with the minimum norm is usually considered: the minimum-norm least-squares solution is given

by

$$\mathbf{x}^+ = A^+\mathbf{b}, \tag{2.15}$$

where A^+ is the **Moore-Penrose inverse** of A. The Moore-Penrose inverse is a type of generalised inverse, a concept that was introduced for matrices that do not have an inverse due to their being singular or rectangular. The Moore-Penrose inverse is uniquely defined by the equations [BIG74, Bar90]

$$\begin{aligned} AA^+A &= A & (AA^+)^H &= AA^+ \\ A^+AA^+ &= A^+ & (A^+A)^H &= A^+A, \end{aligned} \tag{2.16}$$

and the following properties may be deduced: [Bar90]

$$\begin{aligned} (A^+)^+ &= A & (A^HA)^+ &= A^+(A^H)^+ \\ (A^H)^+ &= (A^+)^H & (AA^H)^+ &= (A^H)^+A^+. \end{aligned} \tag{2.17}$$

Note that in addition to the Moore-Penrose inverse, there exists a range of generalised inverses that satisfy some but not all of the equations in Eq. (2.16). The Moore-Penrose inverse, also called the pseudo-inverse, may be computed using SVD:

$$A^+ = V\Sigma^+U^H, \tag{2.18}$$

where Σ^+ is an $n \times m$ matrix with the reciprocals of the (non-zero) singular values on its diagonal and zeros elsewhere; that is, the diagonal elements are

$$1/\sigma_1, 1/\sigma_2, \ldots, 1/\sigma_r, \underbrace{0, 0, \ldots, 0}_{\min(m,n)-r}. \tag{2.19}$$

2.3.3 Special matrices

The **Fourier matrix** is used extensively in discrete Fourier analysis to represent the DFT in Eq. (2.3). The Fourier matrix for 1-D sequences of size N, W_N, is $N \times N$ and has elements

$$(W_N)_{kl} = \frac{1}{\sqrt{N}} w_N^{-kl}, \tag{2.20}$$

where $w_N = \exp(i2\pi/N)$ and $k, l = 0, 1, \ldots, N-1$. The Fourier matrix for M-D sequences, $W_\mathbf{N}$, may be obtained by the Kronecker product,

$$W_\mathbf{N} = W_{N_{M-1}} \otimes W_{N_{M-2}} \otimes \cdots \otimes W_{N_0}, \tag{2.21}$$

and its elements are

$$(W_{\mathbf{N}})_{kl} = \prod_{j=0}^{M-1} \frac{1}{\sqrt{N_j}} w_{N_j}^{-(\mathcal{S}^{-1}(k,\mathbf{N}))_j (\mathcal{S}^{-1}(l,\mathbf{N}))_j}, \tag{2.22}$$

where $w_{N_j} = \exp(i2\pi/N_j)$ and $k, l = 0, 1, \ldots, N_0 N_1 \cdots N_{M-1} - 1$. Note that the Fourier matrix is unitary, i.e. $W_{\mathbf{N}}^{-1} = W_{\mathbf{N}}^H$; $W_{\mathbf{N}}^{-1}$ is referred to as the inverse Fourier matrix.

The DFT of an M-D sequence, $f[\mathbf{n}]$, may be computed by stacking the sequence (according to the convention discussed in Section 2.3.1) into a vector, $\mathbf{f}$, and pre-multiplying by the Fourier matrix: i.e. $W_{\mathbf{N}}\mathbf{f}$. Similarly, the IDFT may be computed by pre-multiplying by the inverse Fourier matrix.

An $N \times N$ **circulant matrix** has the form

$$C = \begin{bmatrix} c_0 & c_1 & c_2 & \cdots & c_{N-1} \\ c_{N-1} & c_0 & c_1 & \cdots & c_{N-2} \\ c_{N-2} & c_{N-1} & c_0 & \cdots & c_{N-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_1 & c_2 & c_3 & \cdots & c_0 \end{bmatrix}. \tag{2.23}$$

It is clear that a circulant matrix (or simply 'circulant') is completely defined by its first row, leading to the notation $C = \mathrm{circ}(\mathbf{c})$ where $\mathbf{c} = [c_0, c_1, \ldots, c_{N-1}]^T$. Circulants arise in problems of 1-D circular convolution or correlation and are diagonalised by the Fourier matrix, where the diagonal elements are simply the DFT of the first row of C. Sometimes it may be preferable to consider the circulant whose first column is $\mathbf{c}$, i.e. C^T (pre-multiplication of a vector by a circulant whose first row or column is $\mathbf{c}$ is equivalent to 1-D circular correlation or convolution, respectively, with $\mathbf{c}$). These relationships are summarised as follows:

$$\begin{aligned} W_N^{-1} C W_N &= \sqrt{N}\,\mathrm{diag}(W_N \mathbf{c}) & \qquad W_N C^T W_N^{-1} &= \sqrt{N}\,\mathrm{diag}(W_N \mathbf{c}) \\ W_N C W_N^{-1} &= \sqrt{N}\,\mathrm{diag}(W_N^{-1} \mathbf{c}) & \qquad W_N^{-1} C^T W_N &= \sqrt{N}\,\mathrm{diag}(W_N^{-1} \mathbf{c}). \end{aligned} \tag{2.24}$$

If the elements c_i were instead $L \times L$ matrices, then C would be a block circulant matrix, and if the $L \times L$ matrices were themselves circulant, C would be block circulant with circulant blocks (BCCB); in both cases C would be $LN \times LN$. These matrices arise in 2-D convolution problems, for example. In general, if an M-D sequence $f[\mathbf{n}]$ is stacked into a vector $\mathbf{f}$, then circular convolution with $\mathbf{f}$ is achieved through pre-multiplication by the M-D circulant matrix

$$C = \sqrt{N}\, W_{\mathbf{N}}^{-1}\, \mathrm{diag}(W_{\mathbf{N}}\mathbf{f}) W_{\mathbf{N}}, \tag{2.25}$$

where $N = N_0 N_1 \cdots N_{M-1}$. Multiplication of a vector by an M-D circulant matrix may be achieved quickly via M-D FFTs: in 1-D for example, if $C = \mathrm{circ}(\mathbf{c})^T$ then

$$C\mathbf{x} = \sqrt{N}\, W_N^{-1}[(W_N \mathbf{c}) \circ (W_N \mathbf{x})]. \tag{2.26}$$

Ref. [Dav79] is the standard reference for circulant matrices.

A circulant is a special case of a **Toeplitz matrix**, which has equal elements along diagonals parallel to the principal diagonal. Block Toeplitz and block Toeplitz with Toeplitz blocks (BTTB) matrices may also be defined analogously to circulants. Importantly, structured matrices such as these can be inverted using much faster algorithms than are available for an arbitrary matrix. Multiplication of a vector by a Toeplitz matrix of order N can be achieved quickly: the Toeplitz matrix is completed to a circulant of order $2N - 1$ (i.e. the original Toeplitz matrix is embedded in a larger circulant), the vector is completed to length $2N - 1$ by adding $N - 1$ zero elements, and the last $N - 1$ components from the result of fast circulant multiplication are dropped. Multiplication of a vector by an M-D Toeplitz matrix (i.e. a BTTB matrix in 2-D) can be performed analogously by fast M-D circulant multiplication.

An $N \times N$ **permutation matrix**, P, is the identity matrix, I, with rows rearranged according to some permutation. Pre- and post-multiplication by P has the effect of rearranging the rows and columns of a matrix, respectively, according to that permutation. Note that $P^T P = P P^T = I$, so P is orthogonal (i.e. $P^{-1} = P^T$).

2.3.4 Conditioning

A problem is **well-conditioned** if a small change in the parameters (e.g. due to noisy measurements) produces a small change in the solution, and conversely a small change in the parameters for an **ill-conditioned** problem produces a large change in the solution. A common measure of conditioning is the condition number, $\mathrm{cond}(A)$, which bounds the relative change in the solution for a relative perturbation in measurements. The 2-norm condition number (used in the thesis) is equal to the ratio of maximum to minimum singular values,

$$\mathrm{cond}(A) = \frac{\sigma_1}{\sigma_{\min(m,n)}}. \tag{2.27}$$

A more in-depth study of conditioning can be found in Section 4.4.1, where an alternative measure is developed.

2.4 Number theory

Gauss famously described mathematics as the "queen of the sciences" and number theory as the "queen of mathematics". While perhaps the most abstract branch of mathematics, number theory can provide insight into arithmetic involving the Fourier matrix due to its inherently modulo nature: note that $w_N^x = w_N^{(x \bmod N)}$ for $x \in \mathbb{Z}$. The relevant theory is discussed here, while fuller treatments can be found in Refs. [HW56, Red96].

Definition 2.1. A non-zero integer a *divides* an integer b, written $a \mid b$, if there exists an integer c such that $ac = b$. The case where a does not divide b is written $a \nmid b$.

Definition 2.2. An integer g is a *common divisor* of integers a and b if $g \mid a$ and $g \mid b$. The *greatest common divisor* is denoted (a, b). Analogously, the greatest common divisor of integers $a_1, \ldots, a_n$ is denoted $(a_1, \ldots, a_n)$.

Definition 2.3. If the greatest common divisor of integers a and b is one (i.e. $(a, b) = 1$), then a and b are *relatively prime*. Analogously, if the greatest common divisor of integers $a_1, \ldots, a_n$ is one, the integers are *pairwise relatively prime*.

Definition 2.4. For integers a and b and a positive integer m, a is *congruent* to b modulo m, written $a \equiv b \pmod{m}$, if and only if $m \mid (a - b)$.

Theorem 2.3 (Congruences). *Let m be a positive integer.*

(a) If $\{a_1, \ldots, a_n\}$, $\{b_1, \ldots, b_n\}$, and $\{k_1, \ldots, k_n\}$ are any sets of integers such that $a_i \equiv b_i \pmod{m}$, $1 \leq i \leq n$, then

$$\sum_{i=1}^{n} k_i a_i \equiv \sum_{i=1}^{n} k_i b_i \pmod{m}. \tag{2.28}$$

(b) If a, b, c, and d are any integers, then

$$a \equiv b \pmod{m} \text{ and } c \equiv d \pmod{m} \Rightarrow ac \equiv bd \pmod{m}. \tag{2.29}$$

Theorem 2.4 (Chinese Remainder Theorem). *Let $m_1, m_2, \ldots, m_n$ be positive integers that are relatively prime, and let $a_1, a_2, \ldots, a_n$ be any n integers. Then the system of congruences*

$$x \equiv a_1 \pmod{m_1}, x \equiv a_2 \pmod{m_2}, \ldots, x \equiv a_n \pmod{m_n} \tag{2.30}$$

has a unique solution for x modulo $M = m_1 m_2 \cdots m_n$ (i.e. if x_1 and x_2 are both integers that satisfy the system of congruences, then $x_1 \equiv x_2 \pmod{M}$), namely

$$\sum_{i=1}^{n} (M/m_i) b_i a_i, \tag{2.31}$$

where b_i satisfies $(M/m_i) \equiv b_i \pmod{m_i}$, $1 \leq i \leq n$.

2.5 Combinatorics

The finite nature of discrete Fourier analysis provides an application for combinatorial theory, with questions such as: How many ways can p sampling locations be selected from N possible locations? The basics of permutations and combinations are discussed here [Coh78].

Theorem 2.5 (Permutations). *The number of ordered selections of r things out of n distinct objects is*

$$(n)(n-1)\cdots(n-r-1) = \frac{n!}{(n-r)!}. \tag{2.32}$$

Theorem 2.6 (Combinations). *The number of ways in which r objects can be selected out of n without regard to order is*

$$\binom{n}{r} = \frac{n!}{(n-r)!r!} = \frac{(n)(n-1)\cdots(n-r-1)}{(r)(r-1)\cdots(1)}. \tag{2.33}$$

A permutation $a_1, a_2, \ldots, a_n$ of the numbers from 1 to n can be depicted as

$$\begin{pmatrix} 1 & 2 & 3 & \cdots & n \\ a_1 & a_2 & a_3 & \cdots & a_n \end{pmatrix}, \tag{2.34}$$

which means that 1 is mapped to a_1, 2 is mapped to a_2, and so on. Cycle notation is an alternative way of writing permutations, where a cycle is a list of numbers $(a_1, a_2, \ldots, a_m)$ that indicates: a_1 is mapped to a_2, a_2 is mapped to a_3, ..., and a_m is mapped to a_1. The length of a cycle is the number of terms in it. For example,

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 4 & 3 & 6 & 1 & 2 \end{pmatrix} = (1\ 5)(2\ 4\ 6)(3) \tag{2.35}$$

is the product of three cycles, with lengths two, three, and one, respectively.

Consider the following problem: a set of n objects, say the numbers 1 to n, are to be painted with one of m colours. There is a group, G, of permutations of the numbers 1 to n that sends some colourings onto other numbers. A colour r is assigned a weight $w(r)$. Let $|G|$ be the number of permutations in G, and for every permutation $g \in G$, let $\lambda_c(g)$ represent the number of cycles of length c in permutation g. Two colourings of the n objects are considered the same if some permutation $g \in G$ sends one colouring onto the other. The whole class of equivalent colourings of the objects is called a scheme. The pattern inventory, due to Pólya, gives the number of distinct colourings for each scheme [dB64].

Theorem 2.7 (Pólya). *The pattern inventory is*

$$\frac{1}{|G|}\sum_{g\in G} w_1^{\lambda_1(g)} w_2^{\lambda_2(g)} \cdots w_n^{\lambda_n(g)}, \tag{2.36}$$

where $w_k = w(1)^k + w(2)^k + \cdots + w(m)^k$, *and* $k = 1, 2, \ldots, n$.

2.6 Inverse problems

In the physical world, "a **direct problem** is a problem oriented along a cause-effect sequence or, also, a problem which consists of computing the consequences of given causes; then, the corresponding **inverse problem** is associated with the reversal of the cause-effect sequence and consists of finding the unknown causes of known consequences" [BB98]. In MR imaging, the direct problem is computing the MR signals produced given knowledge of the magnetic properties of the object being imaged, while the inverse problem involves determining these properties from the MR signals, given knowledge of how the signals were produced.

A problem is **well-posed** when its solution is unique, exists for arbitrary data, and depends continuously on the data. Typically, however, inverse problems are **ill-posed** due to a loss of information inherent in the direct problem. If the system is linear, the general relationship between an object, $f(t)$, and its image, $g(x)$, is described by the Fredholm integral equation of the first kind: [PMBdM84]

$$g(x) = \int K(x,t) f(t)\, dt. \tag{2.37}$$

If the operator K is integrable, the Riemann-Lebesgue lemma shows that

$$\lim_{x\to\infty} \int K(x,t) \sin xt \; dt \to 0, \tag{2.38}$$

which means small changes in the data may produce large oscillations in the solution.

Since practical problems have discrete data, discretisation is required to produce a linear algebraic system to solve numerically. However, the finer the discretisation of an ill-posed problem, the larger the condition number of the corresponding discrete problem [BB98]. An accurate discretisation, therefore, usually results in an ill-conditioned problem.

The cure for ill-posed problems is to search for approximate solutions and to incorporate additional constraints, using information arising from the physics of the problem that is known *a priori*. For example, the object's energy may be assumed to have an upper bound, or the object itself may be assumed to be smooth, non-negative, or to lie within a bounded

region. These constraints can be included via regularisation methods, a term introduced by the Russian mathematician Tikhonov [Tik63]. A statistical approach, on the other hand, yields so-called Bayesian methods, where the object is assumed to be the realisation of a random process with a known probability distribution.

2.6.1 Solution methods

A variety of methods exist for solving inverse problems. Once the problem is expressed in linear algebraic form, techniques such as least-squares and minimum-norm least-squares (discussed in Section 2.3.2) may be used. Consider the system

$$A\mathbf{f} + \mathbf{n} = \mathbf{g}, \tag{2.39}$$

where $A \in \mathbb{C}^{m \times n}$ and $\mathbf{f}$, $\mathbf{n}$, and $\mathbf{g}$ are the object, noise, and (noisy) image, respectively. Assuming that $m \geq n$, the solution via pseudo-inverse may be written as

$$\mathbf{f}^+ = A^+\mathbf{g} = VS^+U^H\mathbf{g} = \sum_{i=1}^{n} \frac{1}{\sigma_i}(\mathbf{u}_i^H\mathbf{g})\mathbf{v}_i, \tag{2.40}$$

where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the ith columns of U and V, respectively. If the image is expressed as the sum of the noise-free image, $\mathbf{g}_0$, and the noise, $\mathbf{g} = \mathbf{g}_0 + \mathbf{n}$, this becomes

$$\mathbf{f}^+ = \sum_{i=1}^{n} \frac{1}{\sigma_i}(\mathbf{u}_i^H\mathbf{g}_0)\mathbf{v}_i + \sum_{i=1}^{n} \frac{1}{\sigma_i}(\mathbf{u}_i^H\mathbf{n})\mathbf{v}_i. \tag{2.41}$$

Small values of σ_i therefore amplify the noise: by eliminating the singular vectors with small singular values from the reconstruction, much of the error due to noise can be eliminated [WN94]. In this case, the reconstruction is simply truncated to achieve regularisation. Alternatively, incorporating an energy constraint gives Tikhonov-Miller regularisation [Mil70], where the singular values are weighted according to the factor $\sigma_i/(\sigma_i^2 + \alpha^2)$ which varies according to the parameter α.

Iterative solutions may also be employed to solve the system, such as steepest descent or the conjugate gradient method [Lue84, She94b]. Iterative methods have the property of 'semi-convergence', where they converge to the correct solution initially but then diverge, due to propagation of the noise dominating the error; early termination is therefore used as a regularisation technique. The conjugate gradient method is much faster than steepest descent, particularly if preconditioning is used, and it also has inherent regularisation [BB98]. The method of projection onto convex sets [SS87] is another iterative technique that may be used.

If the object can be modelled as a realisation of a random process with a known prob-

ability distribution, Bayesian methods may be used. For example, in MR images (and others) the histogram of differences of intensity of nearest neighbour pixels has Lorentzian shape [WSB+01]. Maximum *a posteriori* (MAP) techniques attempt to find the most probable image, $\hat{\mathbf{f}}$, given the data collected, i.e. $\max_{\hat{\mathbf{f}}} P(\hat{\mathbf{f}}|\mathbf{g})$. Using Bayes' theorem,

$$P(\hat{\mathbf{f}}|\mathbf{g}) = \frac{P(\mathbf{g}|\hat{\mathbf{f}})P(\hat{\mathbf{f}})}{P(\mathbf{g})}, \tag{2.42}$$

and maximising $\log P(\hat{\mathbf{f}}|\mathbf{g})$, the problem becomes

$$\max_{\hat{\mathbf{f}}} \left[\log P(\mathbf{g}|\hat{\mathbf{f}}) + \log P(\hat{\mathbf{f}})\right]. \tag{2.43}$$

The first 'likelihood' term gives the probability of the data, given the estimated object, while the second 'prior' term evaluates the probability of the estimated object according to the assumed probability distribution. (Maximum likelihood (ML) techniques use the first term but not the second; it is the incorporation of prior information that distinguishes MAP from ML.) The optimisation problem can be solved by techniques such as the conjugate gradient method, but they are typically quite slow, due predominantly to calculation of the likelihood term.

2.6.2 Consequences for the MR inverse problem

In Section 1.2, three questions were posed for the research:

1. How few samples can be measured while still enabling a good image to be formed?
2. Where should the samples be placed?
3. What reconstruction algorithm should be used?

The first question immediately poses a further question: What constitutes a 'good' image? Clearly the image should be 'close' to the true object in some sense, but determining 'how close' is an interesting problem in itself. At this stage, it suffices to recognise that some definition of image quality is required.

In the present context of inverse problems, the issues involved in these questions may be elucidated. Consider the system in Eq. (2.39), where an estimate of the object, $\hat{\mathbf{f}}$, is desired. Questions one and two concern the number and locations of samples, which directly affect the system matrix, A, thereby influencing its properties. For example, if the system is under-determined, exact reconstruction of every point in the object ('full reconstruction') will be impossible even in the absence of noise. However, even an exactly determined or over-determined system does not guarantee full reconstruction: A must have full column rank to make this possible. In the noisy environment that accompanies real-world sam-

pling, the singular values of A determine the sensitivity of the reconstruction. The sample-selection process is therefore concerned with choosing a matrix A that has full column rank and is well-conditioned.

Question three concerns the reconstruction technique used, which has an impact on the estimate obtained, $\hat{\mathbf{f}}$. At first glance, how close the estimate is to the true object would seem to be the sole criterion by which to judge alternative techniques. However, linear algebraic techniques may become intractable for large images due to the very large matrices involved, while iterative solutions may be very slow to converge. In addition, some methods may be less sensitive to noise than others. Evaluating a reconstruction technique, therefore, must take the following into account: accuracy, practicality, speed, and sensitivity to noise.

Chapter 3

Magnetic resonance imaging

> The variations in water contents and proton relaxation times among biological tissues should permit the generation . . . of useful . . . images from the rather sharp water resonances of organisms, selectively picturing the various soft structures and tissues.
>
> Paul C. Lauterbur, *Nature*, 1973

Wilhelm Röntgen literally changed the way we see the world when he discovered x-rays in 1895. Medicine was revolutionised by the ability to look inside a patient's body without surgical intervention, as illustrated by the fact that clinical use began as early as three months after Röntgen's first attempts. Today a variety of imaging techniques exist, including computed tomography (CT), ultrasound, positron emission tomography (PET), and magnetic resonance imaging (MRI). These may all be grouped under the term tomography, defined as the process of "creating images of the internal (anatomical or functional) organisation of an object"[1] [LL00]. The advent of the digital computer aided development of tomographic techniques, resulting in the creation of the first CT scanner in 1972 [Hou72].

The phenomenon of nuclear magnetic resonance (NMR) was well-known in chemistry and physics prior to the 1970s, after being independently observed by Felix Bloch and Edward Purcell in 1946, but it was Paul Lauterbur who first showed its potential for imaging in his 1973 article [Lau73]. The word 'nuclear' was later omitted in the imaging context, mainly to allay public fears about safety—unlike CT, which uses x-rays, MR does not use ionising radiation and is not known to present any health risks under normal operating conditions.[2]

[1] Tomography is derived from the Greek *tomos*, meaning 'cut', and is often defined as reconstruction from projections, or imaging of an object slice. The more general definition used here has increased in usage, and is given as an alternative on `www.dictionary.com`: "(medicine) obtaining pictures of the interior of the body".

[2] It should be noted that due to the strong magnetic fields involved, patients with metallic implants (such as artificial joints or a pacemaker) cannot be imaged. In addition, there are several possible sources of biological

The field of MRI has grown substantially since its inception and its applications are now diverse, ranging from static high-resolution imaging of a malignant tumour, to cinematic imaging of a beating heart, and to functional imaging showing active areas of the brain. This chapter gives a brief account of the fundamentals of MR imaging—the practical application for the methods described in the rest of the thesis. More comprehensive treatments can be found in Refs. [CS92, Cal93, Nis95, LL00].

3.1 Nuclear magnetic resonance

The essential property that makes NMR possible is that atoms with an odd number of protons and/or neutrons possess a nuclear spin angular momentum (usually simply called 'spin'). While this is the domain of quantum mechanics, in classical terms a sub-atomic particle can be pictured as spinning around its own axis like a spinning top, giving rise to a small magnetic dipole moment. For biological studies, the hydrogen atom (^{1}H) is the most important: in the form of water (H_2O), it is the most abundant and also gives rise to the largest signals. Its nucleus comprises only one proton, giving it a spin of $+\frac{1}{2}$. Other atoms of interest include phosphorus (^{31}P), which is an important indicator of metabolism.

A large number of these microscopic magnetic moments produce a net magnetic moment at the macroscopic level, sometimes called 'bulk magnetisation'. Ordinarily, the spins have random orientations due to thermal noise, creating a net magnetic moment of zero. However, if an external magnetic field, B_0, is applied, the spins become aligned in the direction of the field and create a non-zero net magnetic moment. The direction of B_0 is called the z-direction or longitudinal direction by convention. In addition, the field causes the atoms to exhibit resonance at the *Larmor* frequency,

$$\omega_0 = \gamma B_0, \tag{3.1}$$

where γ is the gyromagnetic ratio, a unique characteristic of a particular atom. A related constant, $\gamma\!\!\!^{-} = \gamma/2\pi$, is commonly used. For ^{1}H, $\gamma\!\!\!^{-} = 42.58$MHz/T, so at 1.5T (a standard magnetic field strength in clinical use) the hydrogen atoms resonate at 63.87MHz. Table 3.1 presents the gyromagnetic ratios and relative sensitivities for some relevant nuclei [Nis95].

To produce an NMR signal, a radiofrequency (RF) electromagnetic field, B_1, is set at the Larmor frequency and applied to the system by a short RF pulse. The field is oriented perpendicular to the main magnetic field, in the x-y or transverse plane. The interaction of the resonating spinning magnetisation and the resonating alternating magnetic field (a driving force) induces a slight rotation of the axis of net magnetisation. After this perturbation

hazards, including strong static magnetic fields, alternating electric and magnetic fields, MR contrast agents, and loud noise due to rapidly switching magnetic fields. However, under normal clinical usage the technique is generally viewed as a safe procedure.

Table 3.1 Gyromagnetic ratios and relative sensitivities for relevant nuclei

nucleus	spin	gyromagnetic ratio γ (MHz/T)	relative sensitivity
^{1}H	1/2	42.575	1.000
^{13}C	1/2	10.705	0.016
^{19}F	1/2	40.054	0.830
^{23}Na	3/2	11.262	0.093
^{31}P	1/2	17.235	0.066

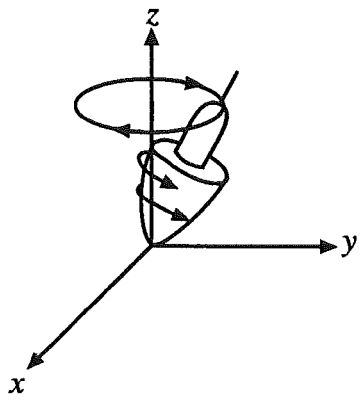


Figure 3.1 The precession of a spinning top under gravity is similar to the precession of nuclear spin after application of an RF field [LL00].

from the main magnetic field, the net magnetisation will *precess* at the Larmor frequency; that is, the net magnetisation will rotate about the z-axis at a well-defined frequency. In classical terms, this can be visualised as a spinning top precessing under the external force of gravity, as depicted in Fig. 3.1. The angle that the net magnetisation is rotated off the z-axis is known as the *flip angle*, α.[3]

The strength and duration of the RF pulse that produces the RF field determines the flip angle obtained; for example, following a pulse selected to produce a 90° flip angle, all net magnetisation lies in the transverse plane. After the field is switched off, the spins undergo relaxation back to equilibrium. By Faraday's law, the rotating magnetic moments induce an electromotive force (EMF) in an RF receiver coil oriented to detect changes in magnetisation in the transverse plane. The received signal is called a *free induction decay* (FID)[4], and a large number of these are used to produce an MR image.

[3] In quantum terms, spin $+\frac{1}{2}$ particles can only exist in two discrete states: parallel and antiparallel (with respect to the magnetic field). It is only at the macroscopic level that the net magnetisation can be described as having orientations of different angles.

[4] 'Free' refers to the fact that the signal is produced by free precession about the B_0 field; 'induction' signifies the application of Faraday's law of electromagnetic induction; and 'decay' refers to the decrease in signal amplitude over time.

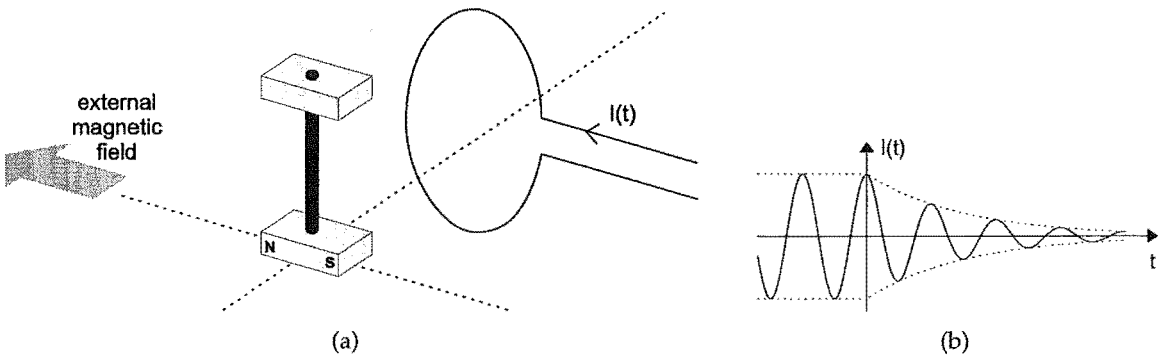


Figure 3.2 A mechanical magnet-and-coil system bears similarities to the situation in NMR.

To give an intuitive feel for the underlying physics, consider the following mechanical NMR analogy [CS92]. The mechanical system shown in Fig. 3.2(a) consists of: a bar magnet suspended by a rigid rod that allows rotation, an external magnetic field, and a coil that is able to produce a field perpendicular to the main field. In equilibrium, the magnet is parallel to the coil. If a DC current were applied to the coil, the bar magnet would be pushed out of equilibrium. If the current were then removed, the magnet would rock back and forth back to equilibrium, inducing a secondary current in the coil at the resonant frequency of the system. If instead an AC current were applied to the coil at the resonant frequency, the oscillations would be enhanced and a free induction decay would again occur once the current was removed. Fig. 3.2(b) depicts the resulting secondary current in the coil if the applied AC current were switched off at $t = 0$.

3.2 Intrinsic MR parameters

The ability of MR to produce useful images is due to differences in the intrinsic MR properties of different tissues. The magnetic characteristics of a particular tissue are determined by its **magnetic susceptibility**, either diamagnetic, paramagnetic, or ferromagnetic. **Proton spin density** measures the abundance of nuclei measured in the MR experiment, but it is not a set physical value, dependent on both experimental and tissue parameters. For example, a calcium object gives no signal in an experiment tailored to ^{1}H, and while water gives a very strong signal in such an experiment, ice gives no signal due to a very short T_2 time (discussed below), since the signal has decayed away before it can be recorded.

After the RF field is switched off, the net magnetic moment gradually realigns with the B_0 field as the spin system returns to equilibrium: the T_1 **relaxation time** (also called longitudinal or spin-lattice relaxation time) measures how rapidly this occurs. The RF excitation pulse deposits energy into the spin system, and the term spin-lattice refers to the fact that as

the system returns to equilibrium, this energy is transferred from the spins into the adjacent molecular lattice.[5] Analogously, a heated object transfers heat energy into the environment as it cools, at a rate dependent on the shape and material of the object. T_1 is a time constant that describes the rate at which the net magnetisation in the longitudinal direction returns to its equilibrium level. Since it is magnetisation in the transverse plane that is measured by the receiver, the realignment of the net magnetic moment with the z-axis corresponds to a decay in signal strength. Values of T_1 are greater at higher field strength and in humans T_1 ranges between about 100–1500ms. Fig. 3.3(a) shows the exponential rise in z-axis net magnetisation governed by T_1.

The signal strength also decays due to alterations in phases ('dephasing') as a result of the spins precessing at different frequencies. The dephasing occurs due to both fixed and time-varying field variations, which in turn affect the resonant frequencies of individual spins and cause phase differences. This loss of coherence results in reduced signal strength, but unlike T_1 decay, it is not due to energy transfer. Sources of fixed field variations include magnetic field inhomogeneities or differences in tissue susceptibilities (see Section 3.5), while time-varying variations can be caused by movement of spins at the microscopic level or interaction of the magnetisation with protons, which are themselves slightly magnetic. The **T_2 relaxation time** (also called transverse or spin-spin relaxation time) measures how quickly the time-varying component of this decay mechanism occurs, while the **T_2^* relaxation time** (pronounced 'tee two star') encompasses both fixed and time-varying effects, meaning T_2 is always longer than T_2^*. T_2 is largely independent of field strength, and in humans it ranges between about 20–300ms. The exponential decay in transverse magnetisation governed by T_2 is shown in Fig. 3.3(b).

The magnetic field that an individual ^{1}H nucleus experiences is affected by the local chemical environment. The electron configuration of a particular molecule results in electron 'shielding', that is, a small magnetic field that tends to oppose the external field. The spins 'see' a combination of the local and external fields, therefore resulting in local field inhomogeneities. Since the Larmor frequency is proportional to field strength, this results in a localised change in frequency called the **chemical shift**. For example, in conventional ^{1}H MR, the signals from water (H_2O) and fat (lipids, many CH_2 groups) are approximately 250Hz apart at 1.5T, or 3.5 parts per million (ppm).

[5] In quantum terms, the RF pulse excites some spins into higher energy levels. For spin $+\frac{1}{2}$ particles, some of those in the parallel state are excited into the antiparallel state, and as they return to parallel, photons of energy are emitted.

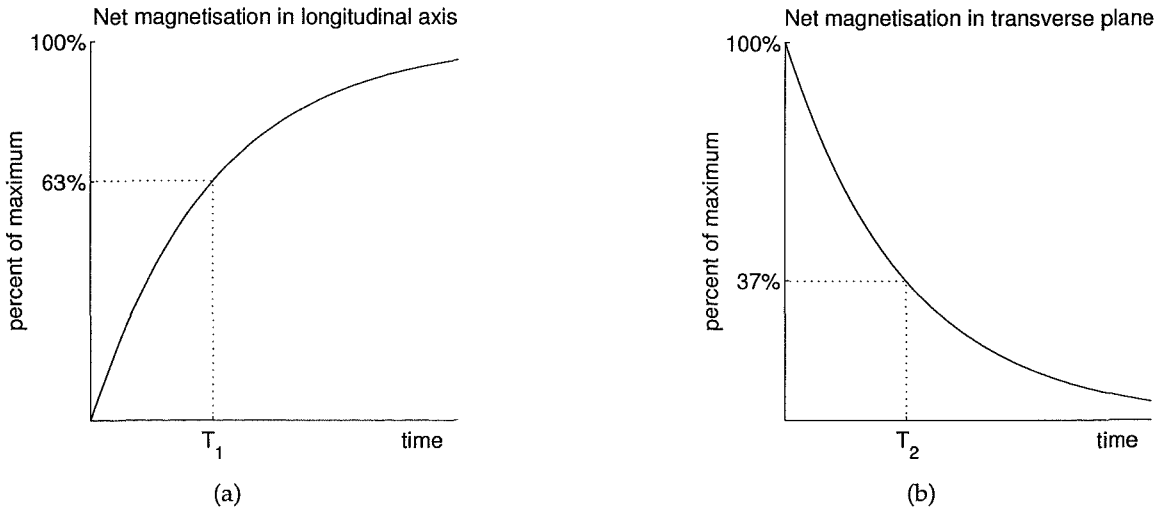


Figure 3.3 The effects of (a) T_1 and (b) T_2 relaxation times on longitudinal and transverse net magnetisation, respectively.

3.3 Signal generation

The simplest form of MR signal is free induction decay (described in Section 3.1), where the FID signal is received directly after application of the RF pulse. Another form of signal is known as *echo*, which differs from FID by providing a two-sided signal, which is important for symmetric coverage of *k*-space (discussed in Section 3.4). An echo signal can be generated either by multiple RF pulses or magnetic field gradient reversal, the former known as **spin echo** signals and the latter as **gradient echo** signals.[6]

A spin echo (SE) signal is formed from two RF pulses: a 90° pulse followed by a 180° pulse after a short time interval, $T_E/2$. Following the initial pulse, the spins are precessing in the transverse plane, but gradually lose phase coherence due to the spins precessing at different rates, causing the received signal to decay according to T_2^*. The 180° pulse then flips the spins to the other side of the transverse plane. Any spins that were leading others are now lagging by the same amount, though still precessing at a higher speed. After another time interval $T_E/2$, the lagging spins will have caught up to the leading spins, recreating phase coherence; the time between the initial RF pulse and the echo is called the *echo time*, T_E. However, the spins will not perfectly align due to time-varying inhomogeneities in the local field, as captured by the T_2 time constant. It is therefore possible to measure T_2 using an SE pulse sequence by examining the level of decay for different values of T_E. Fig. 3.4(a) shows how the signal dephases, is refocussed, and then dephases again. The rephasing is responsible for one side of the echo signal and the subsequent dephasing is responsible for the other side. The envelopes for T_2 and T_2^* decay are also shown.

[6]Callaghan notes that the term 'RF echo' might be more appropriate than the commonly used 'spin echo' [Cal93].

A gradient echo (GE) signal is formed from a single RF pulse with some flip angle, α, followed by the application of a *gradient field*. A gradient in the x-direction (for example), G_x, is a magnetic field that changes linearly in the x-direction and is constant in the y-z plane. The overall field, $B_0 + G_x x$, is entirely oriented in the z-direction but the field strength varies with x location, and therefore the resonant frequency of a particular spin depends on its x location: $\omega(x) = \gamma(B_0 + G_x x) = \omega_0 + \gamma G_x x$. Gradient field strengths are usually about 0.01% of the main field strength and are also used for slice selection and spatial localisation, discussed in the next section.

After application of the RF pulse, a negative x-gradient is turned on, which causes spins at different x-locations to precess at slightly different frequencies, resulting in loss of phase coherence. Once the signal has decayed away, a positive gradient of the same strength is applied, causing the spins to rephase and produce an echo signal. However, any B_0 inhomogeneities will mean the spins do not exactly rephase, and the echo will be weaker than the original, as characterised by T_2^* weighting (as opposed to T_2 in SE). A GE pulse sequence is often used in fast imaging in combination with a low flip angle, but it is more sensitive than SE to field inhomogeneities. Fig. 3.4(b) shows the formation of a GE signal; note the decay envelope governed by T_2^*.

3.4 Image generation

3.4.1 Selective excitation

In a uniform external field, all spins are resonant at the same frequency and the RF pulse excites spins throughout the object. To image only a thin 2-D slice or specific 3-D volume of the object, the first task is **selective excitation** and is achieved using gradient fields in conjunction with a selective RF pulse. As discussed in the previous section, gradient fields alter the overall magnetic field in a particular direction, making the spins' resonant frequencies vary according to location.

For a patient lying supine (face up), slices through the body are called axial (top/bottom), sagittal (left/right), and coronal (front/back), as shown in Fig. 3.5. An axial slice is selectively excited using a z-gradient, G_z, which causes spins near the top of the head to resonate at higher frequencies than spins near the bottom of the head. The next step is to apply an RF pulse with a band of frequencies corresponding to the slice of interest. Ideally, the RF pulse should have a rectangular frequency spectrum, meaning the pulse should be a sinc function in the time domain. Practically, the pulse must be of finite duration and there exist nonlinearities in the spins' behaviour, meaning the waveform should be slightly different, though still sinc-like. In addition, the slice-selection gradient produces a phase shift across the object, which must be corrected by post-excitation rephasing, a neg-

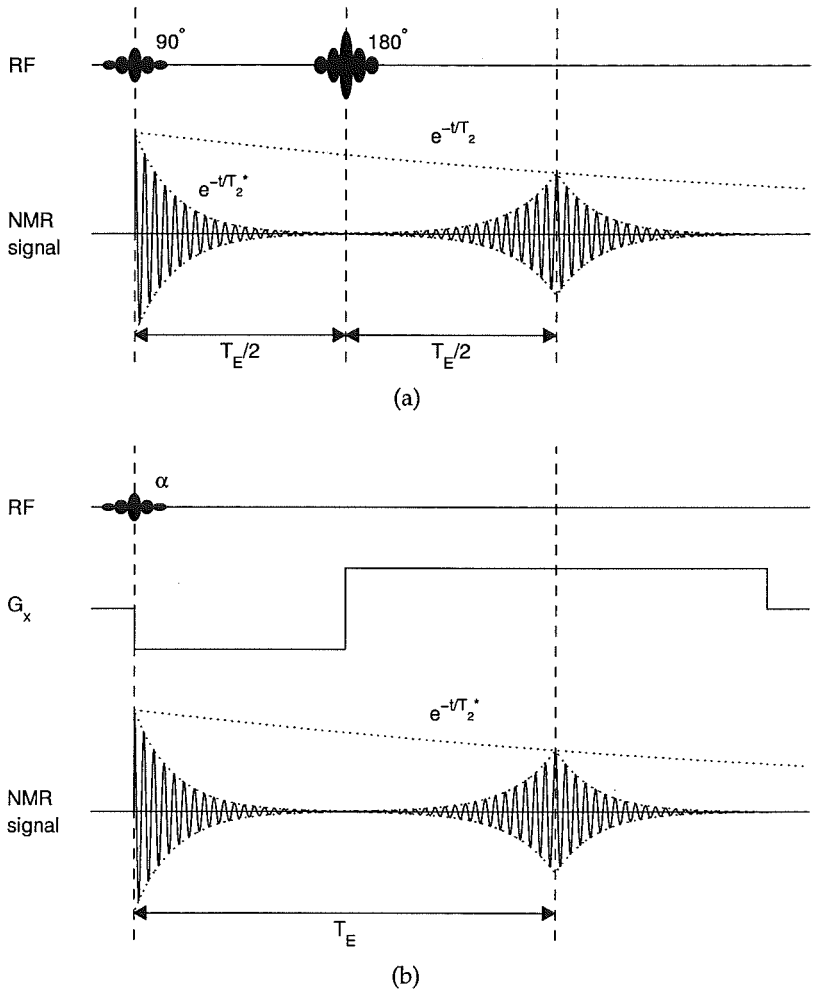


Figure 3.4 Generation of MR signals by (a) spin echo and (b) gradient echo pulse sequences.

ative gradient applied directly following the pulse. Fig. 3.6(a) demonstrates how a band of frequencies results in an excited slice in the presence of a slice-selection gradient, while Fig. 3.6(b) shows the timing of the RF pulse and G_z waveforms.

3.4.2 Spatial information encoding

Following excitation by the RF pulse, spatial information can be encoded into the signal in two ways: **frequency encoding** or **phase encoding**. As already discussed, the presence of a gradient field causes the resonant frequency to vary with spatial location. A frequency-encoding gradient is applied *during* acquisition of the signal, and results in the signal containing a band of frequencies that can be interpreted as different locations within the object. Note that any local variations in resonant frequency are therefore manifested as artefacts in the image (see Section 3.5). Spatial localisation using a fixed frequency-encoding gradient

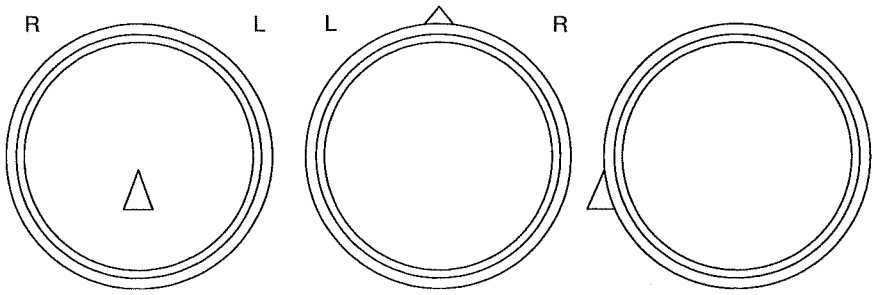


Figure 3.5 Standard section planes through a patient lying supine: (left to right) coronal, axial, and sagittal.

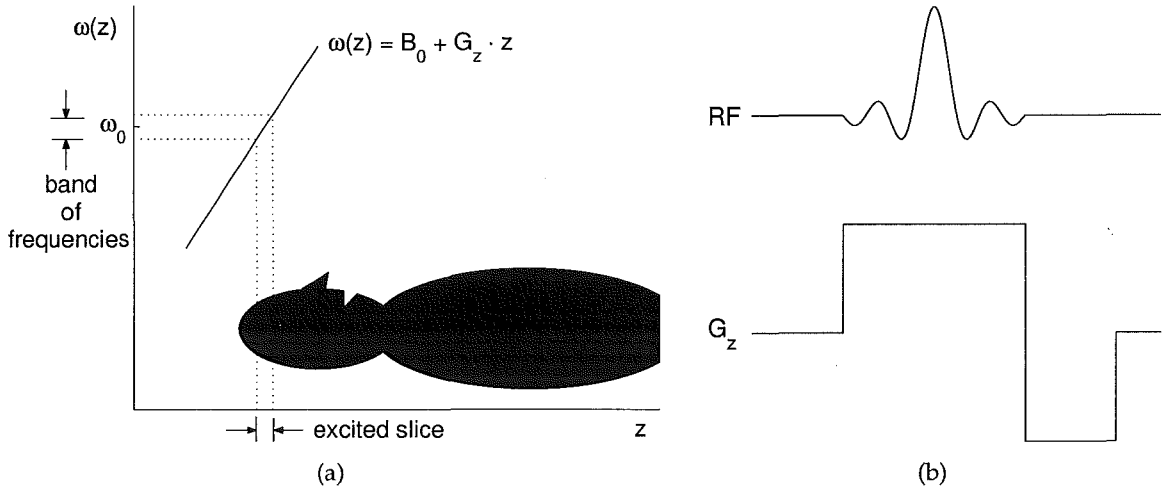


Figure 3.6 Selective excitation: (a) correspondence of a band of frequencies to an excited slice when resonant frequency varies with spatial location, and (b) slice-selective RF pulse and gradient waveforms.

is only possible along one dimension, creating the need for phase encoding for multiple dimensions. A phase-encoding gradient is applied *before* acquisition of the signal; spins precessing at a higher resonant frequency during this time accumulate phase advancement that is dependent on location. Unlike frequency encoding, phase encoding is not a single step process and the signal must be measured at multiple times corresponding to different phase shifts.

3.4.3 Signal equation

The Bloch equation describes the behaviour of the net magnetisation in the presence of a magnetic field (see Ref. [LL00], e.g.), but a simplified model is discussed here [Nis95]. Consider a 2-D object $m(x, y)$, which is a function of the NMR parameters $\rho(x, y)$ (proton spin density), $T_1(x, y)$, and $T_2(x, y)$. The particular pulse sequence used determines the relative influence of each parameter, resulting in an image 'weighting'; an image weighted by one dominant parameter is referred to as T_1-weighted, for example. The spin system

may be envisaged as a collection of oscillators at every location (x, y), with magnitude $m(x, y)$ and phase $\phi(x, y, t)$. The overall signal received, $s_r(t)$, is due to the combined effect of every oscillator,[7] i.e.

$$s_r(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} m(x, y) e^{-i\phi(x,y,t)} dx dy. \tag{3.2}$$

Frequency, $\omega(x, y, t)$, is defined as the rate of change of phase with respect to time, so that

$$\phi(x, y, t) = \int_0^t \omega(x, y, \tau) d\tau = \gamma \int_0^t B(x, y, \tau) d\tau, \tag{3.3}$$

where $B(x, y, t)$ is the overall field, and the Larmor relationship has been used. If the overall field is the sum of the main field and x- and y-direction time-varying gradient fields,

$$B(x, y, t) = B_0 + G_x(t)x + G_y(t)y, \tag{3.4}$$

then the overall phase is given by

$$\begin{aligned} \phi(x, y, t) &= \gamma \int_0^t B_0 d\tau + \gamma \left[\int_0^t G_x(\tau) d\tau \right] x + \gamma \left[\int_0^t G_y(\tau) d\tau \right] y \\ &= \omega_0 t + 2\pi k_x(t) x + 2\pi k_y(t) y, \end{aligned} \tag{3.5}$$

where

$$\begin{aligned} k_x(t) &= \frac{\gamma}{2\pi} \int_0^t G_x(\tau) d\tau \\ k_y(t) &= \frac{\gamma}{2\pi} \int_0^t G_y(\tau) d\tau. \end{aligned} \tag{3.6}$$

The received signal is typically demodulated to remove the carrier frequency, resulting in the following MR signal equation:

$$s(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} m(x, y) e^{-i2\pi(k_x(t)x + k_y(t)y)} dx dy. \tag{3.7}$$

Comparison with Eq. (2.1) reveals the most important relationship in MR imaging—the received signal at time t is equal to the 2-D Fourier transform of the object evaluated at location $(k_x(t), k_y(t))$ in the spatial frequency domain, known in the MR community as 'k-space'. Similarly, the signal received from a 3-D object is its 3-D Fourier transform.

[7] This simplified model ignores the effect of T_2 relaxation. The resultant signal decay may be incorporated by an additional exponential term: $\exp[-t/T_2(x, y)]$.

3.4.4 2DFT imaging

Perhaps the most commonly used imaging method is 2-D Fourier transform (2DFT), or 'spin warp', imaging. As the name suggests, it involves collecting data on a 2-D Cartesian grid, allowing use of the FFT for image reconstruction. Although sampling over an infinite extent at the Nyquist density would allow exact representation of the object (Theorem 2.1), the sampling set must clearly be finite in practice. A truncated data set is therefore obtained [LL00], which results in Gibbs ringing artefact.

A spin echo 2DFT pulse sequence and the corresponding k-space trajectory are shown in Figs. 3.7(a) and (b). The timing diagram only shows a single excitation and the subsequent formation of an echo signal. To gather enough data to form an image, multiple echoes are required; the time between pulse sequences is called the *repetition time*, T_R.

The first step is to excite the desired slice using a 90° selective RF pulse in conjunction with a constant field gradient in the z-direction, with appropriate subsequent rephasing. This causes only those spins in the slice of interest to have net magnetisation in the transverse plane (and therefore to contribute to the received signal). Next, a phase-encoding gradient is applied in the y-direction, which corresponds to a vertical shift in k-space. The dotted lines indicate that in different repetition periods, different strength gradients are applied corresponding to different vertical shifts in k-space. At the same time, a positive frequency-encoding gradient is applied which produces a positive horizontal shift in k-space (the net effect is therefore a diagonal shift). Just prior to the 180° pulse, the location in k-space is the top-right corner, but the pulse causes the trajectory to jump to the conjugate location. The spin echo signal then forms and a final frequency-encoding gradient is applied during data acquisition, causing the k-space trajectory to traverse a horizontal line. Data acquisition involves sampling the received signal, as indicated by the analog-to-digital (A/D) waveform.

The three steps in 2DFT imaging are therefore slice selection, phase encoding, and data acquisition. The x- and y-directions (and gradients) are often referred to as the readout and phase-encoding directions, respectively. A pulse sequence and k-space trajectory for gradient echo 2DFT imaging are shown in Figs. 3.7(c) and (d). In comparison with the spin echo sequence, the differences lie in the RF pulse and G_x gradient: only one RF pulse is used with flip angle α, and the spins are dephased by a negative G_x gradient, which is then reversed to allow data acquisition along a line in k-space.

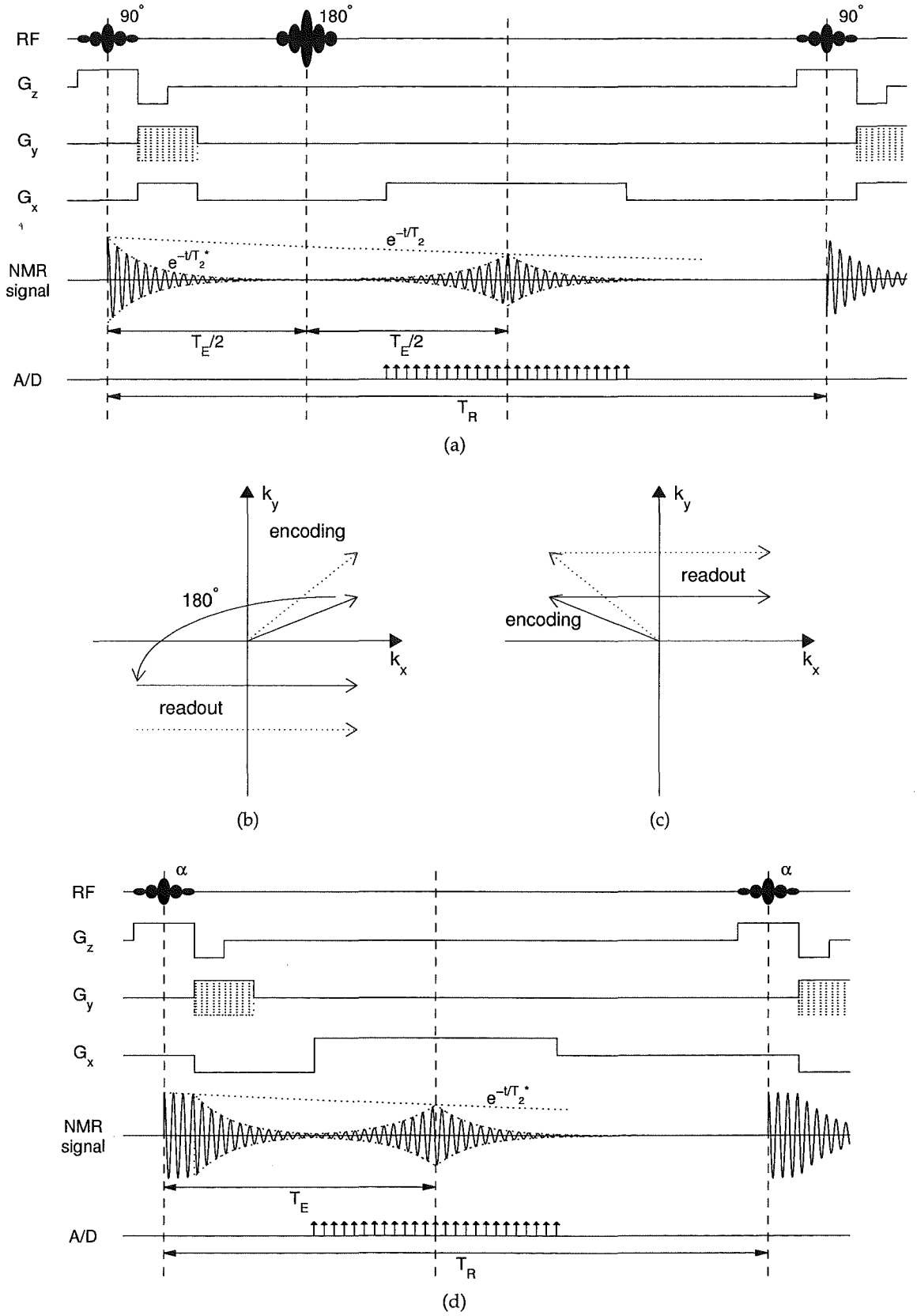


Figure 3.7 2DFT imaging: (a) spin echo pulse sequence, (b) spin echo *k*-space trajectory, (c) gradient echo *k*-space trajectory, and (d) gradient echo pulse sequence.

3.5 Imaging considerations

3.5.1 Non-idealities

With a uniform main field, the resonant frequencies should be identical for all spins throughout the object, although in practice this will not be the case. The consequences of these departures from the ideal resonant frequency are called **off-resonance** effects: first, the B_0 field will not be perfectly homogeneous, although modern scanners incorporate additional 'shim' coils to improve uniformity; second, differences in susceptibility result in local gradient inhomogeneities, such as around the sinuses in the head (i.e. at the boundary of air and tissue); and third, chemical shift occurs due to electron shielding (see Section 3.2). The localised resonance offsets caused by these effects produce consequences such as T_2^* decay due to loss of phase coherence, and result in artefacts and distortion in the image. Since spatial information is encoded with frequency and phase, displacement artefacts can arise; for example, if water and fat were present at the same location in an object, the 2DFT imaging method would produce a displacement between them in the readout direction, with a size dependent on the strength of the readout gradient. In a typical imaging procedure, the displacement might be only one or two pixels, but can be more severe in techniques such as EPI (see Section 3.6.1).

Other non-idealities include T_2 relaxation (which causes signal strength decay, see Footnote 7 above), RF field inhomogeneities (which produce space-varying flip angles), and gradient fields with finite rise times and non-linearities.

3.5.2 Imaging parameters

Apart from the inherent NMR parameters, there are a number of imaging parameters that may be manipulated to influence the image produced, including repetition time (T_R), echo time (T_E), flip angle (α), matrix size (i.e. number of k-space samples), field of view (FOV), and the pulse sequence.

One of the most important measures of image quality is signal-to-noise ratio (SNR), which is defined in the MR community as the ratio of signal amplitude to noise standard deviation. The signal strength depends on the voxel volume ($\Delta x \Delta y \Delta z$), since the larger the voxel, the greater the number of protons contributing to the signal. The main source of noise in MR imaging is thermal and is predominantly due to the receiver coil and the body. The noise is additive, white, and Gaussian, on both in-phase and quadrature received channels (i.e. the real and imaginary parts of the k-space samples). The noise variance is dependent on the bandwidth (BW) of the signal (determined by the gradient field strengths), the

size of the image ($N_x \times N_y$), and the number of signals averaged (N_{avg}). Thus,

$$\text{SNR} \propto (\Delta x \Delta y \Delta z)\sqrt{\frac{N_{\text{avg}} N_x N_y}{\text{BW}}}, \tag{3.8}$$

and since the time between samples of the signal is given by $\Delta t = 1/\text{BW}$,

$$\begin{aligned}\text{SNR} &\propto (\Delta x \Delta y \Delta z)\sqrt{(N_x \Delta t)(N_{\text{avg}} N_y)} \\ &\propto (\text{voxel volume})\sqrt{(\text{A/D time per sequence})(\text{number of sequences})} \\ &\propto (\text{voxel volume})\sqrt{(\text{total measurement time})}.\end{aligned} \tag{3.9}$$

It is therefore the *voxel size* and *total measurement time* that are important. Note that Eq. (3.9) is independent of the size of the image.

SNR considerations have implications for the selection of several imaging parameters. Specifically, in selecting the matrix size, trade-offs must be made between spatial resolution, FOV, and SNR. The consequences for some important parameters are discussed below:

Acquisition time: Longer acquisition times allow for more averaging or more phase encoding steps, increasing the SNR.

Averaging: SNR increases in proportion to the square root of the number of excitations.

Slice thickness: A thicker slice means a larger voxel and therefore more excited protons, but the effects of partial volume averaging (where a tissue occupies only part of the voxel) are exacerbated, lowering spatial resolution.

Field of view: Increasing the FOV increases the voxel size and therefore increases SNR, at the expense of lower spatial resolution.

Spatial resolution: Increasing spatial resolution (without increasing total measurement time) decreases the voxel size, thereby reducing the SNR.

The pulse sequence parameters T_R, T_E, and α can be manipulated to affect the image contrast. After excitation, the net magnetisation undergoes relaxation back to equilibrium at a rate given by T_1; decreasing T_R increases saturation since equilibrium will not have been reached prior to the next excitation, thereby increasing the T_1-weighting of the image (and also decreasing SNR due to reduced signal strength). The signal also decays after excitation due to phase decoherence, as governed by T_2 and T_2^*; increasing T_E increases the dependence on this phenomenon, and therefore increases the T_2-weighting of the image.

Short T_R, short T_E images are therefore called T_1-weighted, while long T_R images are used for proton density and T_2-weighted images. A tissue with longer T_1 will recover less during the T_R period and therefore contribute less to the received signal, making it appear darker (lower amplitude) in T_1-weighted images. Conversely, the signal from a tissue with longer T_2 does not decay as fast, and hence the tissue appears brighter (higher amplitude) in T_2-weighted images. For pulse sequences with relatively short T_R, a 90° pulse results in loss of signal due to saturation of the longitudinal magnetisation. The use of lower flip angles allows sufficient signal strength even with short T_R.

3.6 Fast imaging

In general, imaging time depends on the total amount of data required for a specified spatial resolution and FOV, and the rate at which that data can be collected. Manipulating imaging parameters such as spatial resolution is one way to reduce imaging time, but the aim of fast imaging techniques is to increase speed for a given set of imaging parameters, either by (a) increasing the rate of data collection (fast-scan imaging), or (b) decreasing the amount of data required (reduced-scan imaging).

3.6.1 Fast-scan imaging

A conventional spin echo pulse sequence produces one echo per 90° excitation RF pulse, refocussed by a 180° pulse. The **fast spin echo** (FSE) technique uses multiple 180° RF pulses to produce multiple echoes per excitation pulse. If these are encoded differently, multiple lines of *k*-space can be acquired in a single T_R period (e.g. 8, 16, or more in practice), limited by the rate of T_2 decay of the signal.

In **echo planar imaging** (EPI), multiple lines of *k*-space are also acquired per excitation, but only a single RF pulse is used. The gradient echoes are phase encoded, and an entire rectilinear trajectory can be acquired in a single excitation, or 'shot'. Furthermore, since the signal characteristics are governed by T_2^* decay, the acquisition is very fast, allowing many images to be gathered in less than a second. This high speed brings disadvantages too, though, in the form of more stringent hardware requirements for rapidly switching gradient fields and increased sensitivity to off-resonance effects. Artefacts due to chemical shift can be reduced by suppressing the fat signal (say) using a specially designed RF pulse. In general, a trade-off between speed and quality can be achieved through multi-shot EPI, where several excitations are used to collect all the data.

Another single-shot technique with lesser demands on gradient hardware is **spiral** imaging [BSAG99]. Sinusoidal gradient fields produce a spiral trajectory through *k*-space, introducing the need for a specialised reconstruction algorithm, the most popular of which

is gridding [O'S85, JMNM91]. The oscillating trajectory makes it inherently motion compensated since the high frequencies are all measured at once (in contrast to 2DFT), and off-resonance effects appear as blurring rather than a shift. Other interesting *k*-space trajectories include **radial** [GN93] and **rosette** [Nol97]; the trajectories in the former are radial lines passing through the origin of *k*-space (as used in CT) and in the latter make patterns like rose petals.

Fast gradient echo imaging methods operate with very short T_R periods. When $T_R \ll T_2$, the spin system reaches a state of equilibrium, known as the steady state. In spoiled GRASS techniques (Gradient Refocussed Acquisition in the Steady State, also called FLASH, Fast Low Angle SHot), the residual transverse magnetisation is destroyed before the next RF pulse using so-called spoiler gradients, leaving only the longitudinal magnetisation in steady state [FHM86]. With appropriate choice of T_R, T_E, and α, T_1-, T_2^*-, and proton density-weighted images can be produced. When T_R is on the order of T_2 and the transverse magnetisation is not spoiled, both longitudinal and transverse magnetisation reach the steady state. This phenomenon was first studied by Carr [Car58] and later used for imaging, yielding SSFP techniques (Steady State Free Precession, also called true FISP, Fast Imaging with Steady-state Precession).

3.6.2 Reduced-scan imaging

If additional information is available *a priori* about the object, it can be used to allow reconstruction from an incomplete set of *k*-space data. The fact that the object being imaged is real (i.e. not complex) is utilised in **half-Fourier imaging**. The spectrum of a real object is conjugate symmetric, so only half of *k*-space is required to fully represent the object. In practice, however, field inhomogeneities and object motion cause the realness condition to be violated, introducing a phase term across the object. This can be accounted for by measuring slightly more than half of *k*-space in order to estimate the phase term from the central symmetric portion.

The distribution of intensity differences between adjacent pixels in an MR image is modelled by a Lorentzian distribution. **Bayesian reconstruction** methods in MR imaging use a Lorentzian prior to compensate for incomplete data [WdBF+98, WSB+01]. Another form of prior information is the object's **support**: if the object does not take up the entire FOV then there are regions that are known *a priori* to be identically zero.

Additional information gained from using multiple RF receiver coils is exploited in the **parallel imaging** techniques of SMASH (SiMultaneous Acquisition of Spatial Harmonics) and SENSE (SENSitivity Encoding) to yield reductions in scan time of a half, a third, or more [SM97, PWSB99]. SMASH operates by using linear combinations of coil sensitivities

to synthesise different spatial harmonics and the signals acquired simultaneously can then be combined to produce several lines of k-space at once. No restrictions on coil configuration or sampling pattern are made in SENSE, which is based on the fact that coil sensitivities have an encoding effect complementary to encoding by gradient fields: knowledge of the sensitivity characteristics of each coil allows aliasing due to undersampling to be eliminated.

In dynamic imaging, the temporal resolution can be increased by speeding up the acquisition of each frame. For example, the keyhole technique involves fully sampling k-space initially but then only measuring the inner portion in subsequent frames [JHM+93, vVBD+93, Hu94]. In UNFOLD (UNaliasing by Fourier-encoding the Overlaps using the temporaL Dimension), information from conventional k-space is transferred to the temporal dimension [MGP99]. Lines of k-space are acquired in a time-interleaved fashion, such as even and odd lines sequentially, causing the aliased component of the image to alternate between positive and negative sign. This changes the aliased component's temporal frequency, meaning the aliasing can be removed by temporal low-pass filter. Furthermore, if only the central region of the FOV changes dynamically, more bandwidth can be allocated to the desired image component.

3.7 Applications

MRI techniques have been developed for many clinical applications. In cardiac imaging, the movement of a beating heart presents an inherent challenge to imaging. To minimise motion artefacts, fast spin echo and gradient echo techniques are used, usually accompanied by the patient holding his or her breath! Special pulse sequences can also be employed to provide specific image contrast: for example, double inversion recovery nulls the blood signal in order to observe anatomy.

In functional MRI (fMRI), the aim is to identify active regions of the brain. Neuronal activity is measured indirectly via blood flow: the BOLD (Blood Oxygen Level Dependent) effect describes the series of events that lead to a relative decrease in deoxyhaemoglobin in the activated area, which causes increased signal intensity in a T_2^*-weighted image. In an fMRI examination, the patient is periodically subjected to stimuli and comparisons can later be made between the images obtained with and without stimuli, indicating active areas of the brain for the given task. Because speed is a priority, EPI is the technique of choice.

Cellular metabolic information may be obtained through MR spectroscopy (MRS), by exploiting the chemical shift property discussed in Section 3.2. While ^{1}H spectroscopy can be performed using standard RF coils, other nuclei (such as ^{31}P) precess at different fre-

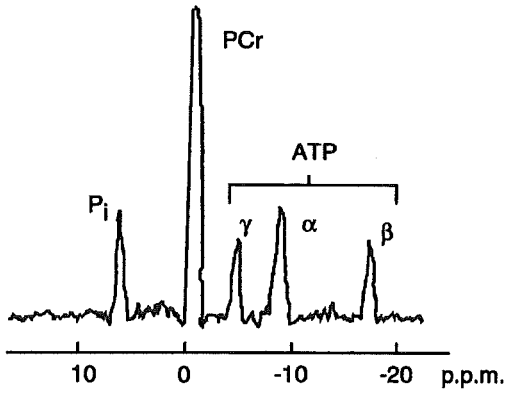


Figure 3.8 Example of a ^{31}P MR spectrum [Nis95].

quencies and require specialised hardware. The name spectroscopy stems from the fact that a *spectrum* is produced: different chemical compounds precess at slightly different frequencies, showing up as peaks in the spectrum of the MR signal obtained. By convention, the frequency axis gives relative shifts in parts per million and decreases from left to right. Fig. 3.8 shows an example of a ^{31}P spectrum: the five characteristic peaks are visible, due to inorganic phosphate (P_i), phosphocreatine (PCr), and adenosine triphosphate (ATP).

As described so far, the spectrum produced is due to tissue throughout the entire excited region of the object. In MR spectroscopic imaging (MRSI), however, a spectrum is produced for every spatial location using phase-encoding gradients (since the chemical shift information is present in frequency deviations, it is not possible to use a frequency-encoding gradient). To form a 2-D spectroscopic image, data is obtained in 3-D k-space: k_x, k_y, and k_f, where the third dimension is the chemical shift axis. A 3-D FFT is then applied to the data, and a 2-D image formed using the magnitude from a particular portion of the spectrum, viz., the portion surrounding the compound of interest. Unfortunately, phase-encoding in both spatial dimensions requires long scan times, usually meaning a trade-off is made that results in poor spatial resolution (e.g. 32×32 pixels).

3.8 Data used for simulations

The Shepp-Logan simulated head phantom [SL74] has become ubiquitous in medical imaging and is used throughout the thesis. It is simply the sum of ten ellipses with parameters listed in Table 3.2 [PK83], and its mathematical nature allows analytic calculation of spectral samples, as shown in Refs. [PK83, VBM+00]. The phantom is shown in Fig. 3.9 for the (usual) FOV defined by the range $[-1, 1]$ in both dimensions. The region of support is simply given by ellipse (a), though a relatively loose support is often assumed in simulations by using an ellipse with the following parameters: centre coordinates $(0, 0)$, major

Table 3.2 Ellipse parameters for the Shepp-Logan head phantom

ellipse	centre coordinates	major axis	minor axis	rotation angle	grey level
a	$(0, 0)$	0.92	0.69	90°	1.0
b	$(0, -0.0184)$	0.874	0.6624	90°	−0.5
c	$(0.22, 0)$	0.31	0.11	72°	−0.2
d	$(-0.22, 0)$	0.41	0.16	108°	−0.2
e	$(0, 0.35)$	0.25	0.21	90°	0.15
f	$(0, 0.1)$	0.046	0.046	0°	0.15
g	$(0, -0.1)$	0.046	0.046	0°	0.15
h	$(-0.08, -0.605)$	0.046	0.023	0°	0.15
i	$(0, -0.605)$	0.023	0.023	0°	0.15
j	$(0.06, -0.605)$	0.046	0.023	90°	0.15

axis 0.98, minor axis 0.75, rotation angle 90°. Use of a relatively loose support is more realistic since the support information may be obtained from 'scout' images, discussed below. In addition, this demonstrates that the methods developed are not restricted to situations where only a tight support is known. Note that the grey levels of the phantom cover the range $[0, 1]$, so root mean square (RMS) error has intuitive meaning. When used in simulations the phantom is sampled to a finite extent in the frequency domain; any Gibbs ringing in the reconstructions is an artefact of the truncated k-space data set and is not due to a reduction in the number of samples.

To simulate measurement noise, spectral 'measurements' of the Shepp-Logan head phantom are corrupted with pseudo-random, complex, white, Gaussian noise as follows: Given a desired SNR, the power in the head phantom is calculated analytically and used to determine the total noise power. The real and imaginary components of the noise are simulated as samples of a Gaussian random variable with zero mean and variance equal to half the total noise power; that is, the noise has uncorrelated real and imaginary parts with equal variance. Mathematically, the total noise power, σ^2_{tot}, is calculated as

$$\sigma^2_{\text{tot}} = \frac{P_{\text{phan}}}{10^{d/10}}, \tag{3.10}$$

where $P_{\text{phan}} = \iint m(x, y)\, dx\, dy$ is the power in the head phantom and d is the desired SNR in decibels. The noise component for a particular spectral sample, n, is then generated by

$$n = x_1 + i\, x_2, \tag{3.11}$$

where x_1 and x_2 are each samples of a Gaussian random variable with zero mean and variance $\sigma^2_{\text{tot}}/2$.

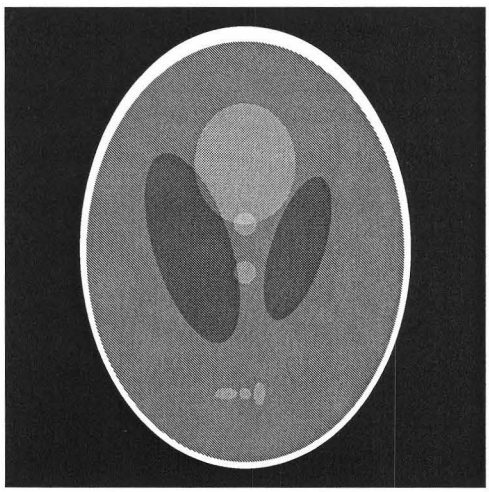


Figure 3.9 The Shepp-Logan head phantom.

In addition to simulated phantoms, *in vivo* data is used to assess the performance of sampling strategies and reconstruction algorithms. A $256 \times 256 \times 128$ 3-D spin warp data set of the author was acquired with a spoiled GRASS pulse sequence (T_R = 22ms, $\alpha = 30°$) on a General Electric Signa 1.5T scanner. The data set is used in several ways: in its entirety, as a single 2-D slice (obtained by computing the IDFT along the longitudinal direction and extracting a slice), as a $240 \times 240 \times 128$ 3-D data set (obtained by truncating the outer parts of k-space; this imitates the situation of this data simply not being measured), and as a 240×240 slice (obtained by truncating the data set and then extracting a slice). In order for RMS error to have intuitive meaning when this data set is used, it is first normalised so that the grey levels of the image cover the range $[0, 1]$.

The prior knowledge used in the methods described in the thesis is restricted to support information. For imaging of the brain, a significant proportion of the FOV contains no signal, thus representing a considerable informational advantage. In most MRI exams, it is standard practice to take preliminary MR images of several axial, coronal, and sagittal slices through the head, known as 'scout' images. In MRSI, a high-resolution MRI is normally acquired prior to spectroscopic imaging. It is envisaged that the support information could easily be gathered from these scout or preliminary MR images via image segmentation. Image processing techniques such as dilation could also be used to create slightly looser supports that may be less vulnerable to small changes in the support, perhaps due to patient movement between the scout images and the full examination.

For the spin warp data set the support information was obtained by analysing the actual images (as opposed to scout images) using the following procedure:

1. threshold the image to form a binary image (using the algorithm in Ref. [Sez90]);
2. perform the morphological operation 'opening' (erosion followed by dilation), repeated until no change occurs;
3. fill any holes that remain in the support.

All numerical computation in this thesis was performed using MATLAB®, whose indexing convention is consistent with that adopted in Section 2.3.1, making the mathematics easy to implement. In simulations, the x, y and z axes are taken to be the first, second and third coordinates, respectively. However, in MATLAB® this means that the x-axis is vertical (matrix rows) and the y-axis is horizontal (matrix columns), so for display purposes the images are transposed to yield the conventional x-y plane. As a consequence of this convention, a discrete sequence of size 120×96 would have 120 elements in the x-direction (horizontal) and 96 elements in the y-direction (vertical). Note that for the spin warp data set, the read-out direction (usually labelled x) is the anterior-posterior axis, but since this is ordinarily displayed on the vertical axis (for an axial slice) it is considered the y-axis.

Chapter 4

Cartesian sampling

4.1 Introduction

In the majority of situations where a signal is sampled, the samples are uniformly spaced. The availability of both the WKS sampling theorem and the FFT, along with the convenience of regularity in many sampling procedures, have no doubt all influenced the prevalence of uniform sampling. For a number of reasons, however, an entire set of regularly spaced samples may be unavailable: some may have been lost or made unusable due to noise for example, or samples may have been deliberately omitted to reduce data acquisition time. Reconstruction is still possible if additional information is available, such as other arbitrarily located samples [Xia95] or prior knowledge of the object's support [Mar86].

Sampling below the Nyquist density was discussed in Section 2.2, with the minimum density given by the Lebesgue measure of the support. In MRI, the support of the object being imaged is simply the region of the body itself. Since this region is generally not an M-D cuboid (i.e. a rectangle in 2-D), sampling at the Nyquist density results in linearly dependent samples, meaning some may be omitted while still allowing full reconstruction. In this chapter, then, sampling strategies are considered that are subsets of the full Nyquist set. While this is ordinarily deemed a special case of irregular sampling, it is referred to in this thesis as Cartesian sampling (CS) to emphasise the fact that samples are located on a regular Cartesian grid.

This approach is applicable to 3-D MRI, MRSI, and other modalities that measure a single point in 2-D k-space per echo. A reduction in the number of echoes required yields a proportional reduction in imaging time. Alternatively, the same imaging time could be used to produce a higher resolution image by covering a greater extent of k-space. A full

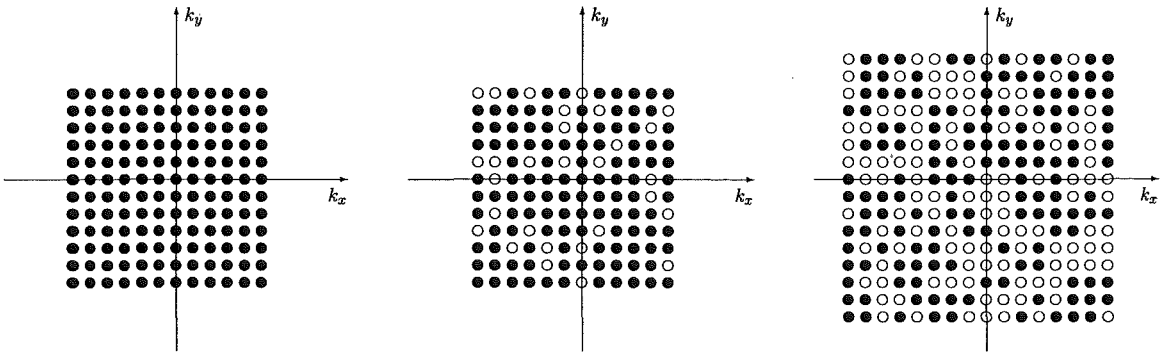


Figure 4.1 Possible implementations of Cartesian sampling in k-space: (a) full grid, (b) 20% reduction in number of samples (and therefore 20% reduction in data acquisition time), and (c) no reduction in the number of samples but higher spatial resolution.

Cartesian grid and examples of these two possibilities are depicted in Fig. 4.1.

This chapter begins by establishing a mathematical framework for the problem before existing reconstruction techniques are discussed. A body of theory is then developed that provides insight into the problem and has practical implications that are exploited in later chapters. Finally, various sampling issues are discussed in light of the theory.

4.1.1 Novel contributions

- **development of the trace metric** (Section 4.4.1)
  This metric appears elsewhere in the literature (e.g. Refs. [RH95, NL99, GR00, VB01]), but the derivation presented here is more rigorous and also includes an analysis of fair comparisons (Theorem 4.1).

- **relationship between 1-D and M-D** (Section 4.4.2)
  The relationship between 1-D and 2-D sequences was asserted in Ref. [BF96], but it is proved here for 1-D and M-D sequences and is described explicitly (Theorem 4.2).

- **equivalent problems** (Section 4.4.3)
  Results on cyclic shifts and mirror images appear in the literature (e.g. Refs. [FB96, GR00]), but the equivalence is proved here (Theorem 4.3), a related new result is obtained (Theorem 4.4), and the number of 'unique' sampling sets is calculated precisely.

- **singular values of the problem** (Section 4.4.4)
  Results are presented on the singular values of the problem, including an application of a previous result (Theorem 4.5), a relatively straightforward property (Theorem 4.6), and an interesting result on complementary problem sizes (Theorem 4.7).

- **contiguous support and maximally spread sampling pattern** (Section 4.4.6)
  A conjecture concerning the optimality of a maximally spread sampling pattern for a contiguous support is presented (Conjecture 4.9) and proved for certain conditions (Theorem 4.10).

- **universal patterns** (Section 4.4.7)
  Results on universal patterns appear in the literature (e.g. Refs. [BF96, VB00]), but an interesting new result on equivalent universal patterns is presented here (Theorem 4.13).

- **spread of sampling locations** (Section 4.5.2)
  A consequence of Conjecture 4.9 is investigated and discussed using an example.

- **conditioning and problem size** (Section 4.5.3)
  It is argued by way of example that larger problems are generally worse conditioned than smaller problems for a similar q/N ratio.

4.2 Mathematical framework

Consider an M-D sampling system where sampling is performed in the spatial frequency domain. (Note that all results are equally applicable if sampling is instead performed in the spatial domain.) Mathematically, this may be described using discrete sequences of size $\mathbf{N}$, as defined in Section 2.1. Let $f[\mathbf{n}]$ be one such sequence, which may be stacked into a vector $\mathbf{f} \in \mathbb{C}^N$, where $N = N_0 N_1 \cdots N_{M-1}$. Similarly, the DFT of $f[\mathbf{n}]$, $F[\mathbf{k}]$, may be represented by $\mathbf{F} \in \mathbb{C}^N$. The relationship between $\mathbf{f}$ and $\mathbf{F}$ is expressed in matrix form (see Section 2.3.3) as

$$W_{\mathbf{N}}\mathbf{f} = \mathbf{F}. \tag{4.1}$$

Assume that the support comprises $q < N$ elements of $\mathbf{f}$; that is, $N - q$ elements at known locations are known *a priori* to be identically zero. Also assume that only $p < N$ elements of $\mathbf{F}$ are in fact measured. Let $Q_{p,N}$ be the set of all strictly increasing integer sequences of length p, with each element of $Q_{p,N}$ being a subset of $\{0, 1, \ldots, N-1\}$. Let $\alpha = \{\alpha_0, \alpha_1, \ldots, \alpha_{p-1}\} \in Q_{p,N}$ and $\beta = \{\beta_0, \beta_1, \ldots, \beta_{q-1}\} \in Q_{q,N}$ denote the indices of the measured samples (sampling set) and non-zero elements (support set), respectively. (Sampling and support sets are often referred to in this thesis as 'patterns'.) As just described, a member of the sampling set, $e \in \alpha$, indexes a location in the vector $\mathbf{F}$ (and analogously for $e \in \beta$ in the vector $\mathbf{f}$). But note that $0 \leq e \leq N-1$, so e uniquely represents a location in the M-D sequence according to the stacking relationship defined in Section 2.3.1. This suggests an alternative representation for sample sets that emphasises the M-D nature of

the underlying sequences: define any set $\boldsymbol{\alpha}$ to have ith component

$$\boldsymbol{\alpha}_i = \mathcal{S}^{-1}(\alpha_i, \mathbf{N}), \quad i = 0, 1, \ldots, |\alpha| - 1. \tag{4.2}$$

The $|\alpha| = |\boldsymbol{\alpha}|$ elements of $\boldsymbol{\alpha}$ are M-D vectors; $(\boldsymbol{\alpha}_i)_j$ denotes the jth component of $\boldsymbol{\alpha}_i$, $j = 0, 1, \ldots, M-1$. The ith element of α is then uniquely expressed by

$$\alpha_i = \mathcal{S}(\boldsymbol{\alpha}_i, \mathbf{N}), \quad i = 0, 1, \ldots, |\boldsymbol{\alpha}| - 1. \tag{4.3}$$

Although the two notations may be used interchangeably, for M-D mathematical expressions the vector notation is favoured (unless $M = 1$ explicitly), while for submatrix indexing and other applications, the scalar form is used.

For brevity, a modification to all set elements is written as a modification to the set itself; for example, the notation $\boldsymbol{\alpha}' = (\boldsymbol{\alpha} + 2) \bmod \mathbf{N}$ is shorthand for

$$\boldsymbol{\alpha}' = \left\{ \mathbf{e} \;\middle|\; \mathbf{e} = (\boldsymbol{\alpha}_i + 2\,\mathbf{1}) \bmod \mathbf{N}, i = 0, 1, \ldots, |\boldsymbol{\alpha}| - 1 \right\}, \tag{4.4}$$

where $\mathbf{1}$ is the all-ones vector, and the modulo function operates element-by-element.

Define the set $Q_{p,\mathbf{N}}$ as

$$Q_{p,\mathbf{N}} = \left\{ \boldsymbol{\alpha} \;\middle|\; \boldsymbol{\alpha} = \mathcal{S}^{-1}(\alpha, \mathbf{N}), \alpha \in Q_{p,N}, N = N_0 N_1 \cdots N_{M-1} \right\}. \tag{4.5}$$

The sampling set and support set can now be designated in the vector notation above as $\boldsymbol{\alpha} \in Q_{p,\mathbf{N}}$ and $\boldsymbol{\beta} \in Q_{q,\mathbf{N}}$, respectively.

By eliminating the unmeasured samples and zero elements from Eq. (4.1), a matrix equation modelling the problem is obtained:

$$W_{\mathbf{N}}[\alpha|\beta]\mathbf{f}[\beta] = \mathbf{F}[\alpha], \tag{4.6}$$

where $W_{\mathbf{N}}[\alpha|\beta]$ is the $p \times q$ submatrix formed from $W_{\mathbf{N}}$ by selecting only rows indexed by α and columns indexed by β. Similarly, $\mathbf{f}[\beta]$ and $\mathbf{F}[\alpha]$ represent $q \times 1$ and $p \times 1$ column subvectors containing the non-zero elements and measured samples, respectively. The term *problem size* is used for the values of $\mathbf{N}$, p and q that parameterise a particular problem. Note that if $p = q$, the symmetry of the Fourier matrix ensures that the problem is entirely equivalent to one in which the sampling and support sets are interchanged. The elements

of $W_{\mathbf{N}}[\alpha|\beta]$ are

$$(W_{\mathbf{N}}[\alpha|\beta])_{kl} = \frac{1}{\sqrt{N}} \prod_{j=0}^{M-1} w_{N_j}^{-(\alpha_k)_j(\beta_l)_j}, \tag{4.7}$$

where $w_{N_j} = \exp(i2\pi/N_j)$, $k = 0, 1, \ldots, p-1$, and $l = 0, 1, \ldots, q-1$. Also note that

$$(W_{\mathbf{N}}[\alpha|\beta]^H W_{\mathbf{N}}[\alpha|\beta])_{kl} = \frac{1}{N} \sum_{i=0}^{p-1} \prod_{j=0}^{M-1} w_{N_j}^{(\alpha_i)_j[(\beta_k)_j-(\beta_l)_j]} \tag{4.8}$$

where $k, l = 0, 1, \ldots, q-1$, and

$$(W_{\mathbf{N}}[\alpha|\beta] W_{\mathbf{N}}[\alpha|\beta]^H)_{kl} = \frac{1}{N} \sum_{i=0}^{q-1} \prod_{j=0}^{M-1} w_{N_j}^{(\beta_i)_j[(\alpha_l)_j-(\alpha_k)_j]}, \tag{4.9}$$

where $k, l = 0, 1, \ldots, p-1$.

It is useful to introduce the following definitions:

Definition 4.1. Let $\boldsymbol{\alpha} \in Q_{p,\mathbf{N}}$. The *complement* of $\boldsymbol{\alpha}$, $\bar{\boldsymbol{\alpha}} \in Q_{N-p,\mathbf{N}}$, is defined as

$$\bar{\boldsymbol{\alpha}} = \left\{\mathcal{S}^{-1}(i, \mathbf{N})\right\}_{i=0}^{N-1} \backslash \boldsymbol{\alpha}. \tag{4.10}$$

Definition 4.2. Let $\boldsymbol{\alpha} \in Q_{p,\mathbf{N}}$. The *mask* of $\boldsymbol{\alpha}$, $H^{\alpha}[\mathbf{k}]$, is a binary sequence with ones determined by the locations in $\boldsymbol{\alpha}$:

$$H^{\alpha}[\mathbf{k}] = \begin{cases} 1 & (\mathbf{k} \bmod \mathbf{N}) \in \boldsymbol{\alpha} \\ 0 & \text{otherwise.} \end{cases} \tag{4.11}$$

The M-D DFT or IDFT of $H^{\alpha}[\mathbf{k}]$ (depending on whether $\boldsymbol{\alpha}$ is a spatial or frequency domain set, respectively) is denoted $h^{\alpha}[\mathbf{n}]$. The mask and its DFT or IDFT may be stacked into vectors as $\mathbf{H}^{\alpha} \in \mathbb{Z}^N$ and $\mathbf{h}^{\alpha} \in \mathbb{C}^N$, respectively.

Definition 4.3. Let $\boldsymbol{\alpha} \in Q_{p,\mathbf{N}}$. The *extraction matrix*, E_{α}, is defined as the $p \times N$ submatrix of the $N \times N$ identity matrix with rows indexed by α, i.e. $E_{\alpha} = I[\alpha|\cdot]$. Note that $E_{\alpha}^T E_{\alpha} = \operatorname{diag}(\mathbf{H}^{\alpha}) \in \mathbb{Z}^{N \times N}$ and $E_{\alpha} E_{\alpha}^T = I \in \mathbb{Z}^{p \times p}$.

Pre-multiplication by E_{α} extracts the rows indexed by α while post-multiplication by E_{α}^T extracts the columns indexed by α. Eq. (4.6) may now be written as

$$(E_{\alpha} W_{\mathbf{N}} E_{\beta}^T)(E_{\beta}\mathbf{f}) = (E_{\alpha}\mathbf{F}). \tag{4.12}$$

4.3 Existing methods

The problem presented above is essentially the same as the traditional problem of superresolution—improving the resolution of a time-domain or spatial-domain object that has been passed through a band-limited system, such as the diffraction limit in an optical system. If the object is limited to a known finite region, then superresolution is theoretically possible [SP61]; the spectrum of a finite object is analytic and any analytic function is completely determined by a finite portion of itself (this is known as analytic continuation). An iterative algorithm for superresolution of continuous functions was developed independently by Gerchberg [Ger74] and Papoulis [Pap75], usually referred to as the Gerchberg-Papoulis (or G-P) algorithm. An extrapolation algorithm for the infinite discrete case was considered in Ref. [JR81], while a finite discrete version of the G-P algorithm was presented in Ref. [Jon86].

Although superresolution usually refers to extrapolation outside the band limit, a broader definition includes interpolation within the band limit [Hun95] and the G-P algorithm can be easily modified to incorporate this case [Fer94]. Clearly then, the G-P algorithm is applicable as a reconstruction technique for the problem of Cartesian sampling. The algorithm proceeds as follows: let the estimate of the spectrum at the kth iteration be $\hat{\mathbf{F}}^k$, and set $\hat{\mathbf{F}}^0 = \mathbf{H}^\alpha \circ \mathbf{F}$, that is

$$(\hat{\mathbf{F}}^0)_i = \begin{cases} (\mathbf{F})_i & i \in \alpha \\ 0 & i \notin \alpha. \end{cases} \tag{4.13}$$

The kth iteration is described by

$$\hat{\mathbf{F}}^{k+1} = \mathbf{H}^{\bar{\alpha}} \circ \{W_{\mathbf{N}}[\mathbf{H}^\beta \circ (W_{\mathbf{N}}^{-1}\hat{\mathbf{F}}^k)]\} + \hat{\mathbf{F}}^0. \tag{4.14}$$

In words, the IDFT of $\hat{\mathbf{F}}^k$ is computed, and the region outside the support is set to zero. The DFT of the result is then computed, and the values at the unmeasured sample locations are placed in the corresponding locations of $\hat{\mathbf{F}}^{k+1}$. The values at the measured sample locations of $\hat{\mathbf{F}}^{k+1}$ are simply set to the measured data. Information is therefore incorporated in both the spatial domain (support) and frequency domain (measurements) at each iteration.

Gerchberg showed that the error (caused by the incomplete measurements) reduces twice per iteration [Ger74]. The algorithm has also been shown to be a form of steepest descent optimisation [CLW95] and a special case of projection onto convex sets (POCS) [SS87]. With noisy measurements, propagation of the noise occurs as the algorithm proceeds, resulting in the phenomenon of semi-convergence (see Section 2.6.1), which may be controlled through early termination or another form of regularisation. With an appropriate convergence criterion, the G-P algorithm can be viewed as a practical, albeit slow [WN94], method for attempting to superresolve realistically sized images. Note that it is straight-

forward to incorporate additional (or alternative) spatial-domain constraints into the algorithm (though it may not then have a decreasing error).

A direct method for superresolution was proposed by Walsh and Nielsen-Delaney [WN94], which amounts to solving Eq. (4.6) by pseudo-inverse with a truncated reconstruction (see Section 2.6.1). They showed that in the absence of noise, the G-P algorithm produces the same result as the direct method, and that in a noisy environment, early termination of the G-P algorithm and truncated reconstruction using their SVD-based method produce very similar results. The difference is due to several factors: the direct method completely eliminates the contributions from particular singular vectors of $W_{\mathrm{N}}[\alpha|\beta]$, while at every G-P iteration, the result contains a partial solution from *every* eigenvector (of a different but related matrix); additionally, performance depends on the distribution of the signal and noise between the singular vectors or eigenvectors for the direct and G-P methods, respectively, and in general these distributions are not identical [WN94]. They also proposed a rule-of-thumb for eliminating singular vectors from the reconstruction: retain those singular vectors corresponding to singular values whose ratio from the maximum singular value is no more than the SNR.

A direct method for regaining missing samples was proposed by Marks [Mar86], applicable to a general uniform sampling geometry. Assuming a Cartesian geometry, the missing samples, $\hat{\mathbf{F}}[\bar{\alpha}]$, are given by

$$\hat{\mathbf{F}}[\bar{\alpha}] = -H_1^{-1} H_2 \mathbf{F}[\alpha], \tag{4.15}$$

where $H_1 \in \mathbb{C}^{(N-p)\times(N-p)}$ and $H_2 \in \mathbb{C}^{(N-p)\times p}$ have elements

$$\begin{aligned} (H_1)_{kl} &= (\mathbf{h}^{\beta})_{S((\bar{\alpha}_k - \bar{\alpha}_l) \bmod \mathbf{N},\mathbf{N})} \quad k,l = 0,1,\ldots,N-p-1 \\ (H_2)_{kl} &= (\mathbf{h}^{\beta})_{S((\bar{\alpha}_k - \alpha_l) \bmod \mathbf{N},\mathbf{N})} \quad k = 0,1,\ldots,N-p-1 \\ &\qquad\qquad\qquad\qquad\quad l = 0,1,\ldots,p-1. \end{aligned} \tag{4.16}$$

The quality of the reconstruction depends on the conditioning of H_1, which depends on the distance among the missing samples, the uniformity of the missing samples' geometry, and the total number of missing samples [CM90].

4.4 Theory

4.4.1 Conditioning

Inspection of the problem formulated in Eq. (4.6) reveals that the performance of a particular sampling set for a given support set is governed by the properties of the matrix $W_{\mathrm{N}}[\alpha|\beta]$. To compare the performance of different sampling sets, it is clearly necessary to have a good measure of conditioning. For an arbitrary matrix, $A \in \mathbb{C}^{m\times n}$, the (2-norm)

condition number, $\mathrm{cond}(A)$, is one measure of performance (see Section 2.3.4), but since it only depends on the maximum and minimum singular values, it often gives an incomplete picture of the underlying behaviour in the presence of noise. Reeves *et al.* used the following criterion for evaluating sampling sets for their selection algorithm: $\mathrm{trace}(A^H A)^{-1}$ if $m \geq n$, and $\mathrm{trace}(AA^H)^{-1}$ if $m < n$ [RH95, GR00]. These were derived from investigating the expected value of the sum of square error in the solution when the measurements are corrupted with zero-mean, unit-variance, additive, white, Gaussian noise. Nagle and Levin also used $\mathrm{trace}(A^H A)^{-1}$ to gauge conditioning [NL99]. A derivation similar to that in Refs. [RH95, GR00] is shown here to develop the *trace metric*, which is used henceforth in the thesis.

The trace metric

Consider the system of Eq. (2.39):

$$A\mathbf{f} + \mathbf{n} = \mathbf{g}, \tag{4.17}$$

where $\mathbf{f}$, $\mathbf{n}$, and $\mathbf{g}$ are the object, noise, and (noisy) image, respectively, and $A \in \mathbb{C}^{p \times q}$ models the imaging process. The minimum-norm least-squares solution is given by $\mathbf{f}^+ = A^+\mathbf{g}$, so the sum of square error (i.e. the error power) is

$$\begin{aligned}
\|\mathbf{e}\|^2 &= \left\|\mathbf{f} - \mathbf{f}^+\right\|^2 \\
&= \left\|\mathbf{f} - A^+\mathbf{g}\right\|^2 \\
&= \left\|\mathbf{f} - A^+(A\mathbf{f} + \mathbf{n})\right\|^2 \\
&= \mathbf{f}^H\mathbf{f} - \mathbf{f}^H(A^+A)\mathbf{f} - \mathbf{f}^H A^+\mathbf{n} \\
&\quad - \mathbf{f}^H(A^+A)^H\mathbf{f} + \mathbf{f}^H(A^+A)^H(A^+A)\mathbf{f} + \mathbf{f}^H(A^+A)^H A^+\mathbf{n} \\
&\quad - \mathbf{n}^H(A^+)^H\mathbf{f} + \mathbf{n}^H(A^+)^H A^+A\mathbf{f} + \mathbf{n}^H(A^+)^H A^+\mathbf{n}.
\end{aligned} \tag{4.18}$$

Using the identities and properties of the pseudo-inverse (Eqs. (2.16) and (2.17)), this can be simplified to

$$\|\mathbf{e}\|^2 = \|\mathbf{e}_r\|^2 + \|\mathbf{e}_n\|^2, \tag{4.19}$$

where

$$\begin{aligned}
\|\mathbf{e}_r\|^2 &= \mathbf{f}^H(I - A^+A)\mathbf{f} \\
\|\mathbf{e}_n\|^2 &= \mathbf{n}^H(AA^H)^+\mathbf{n}.
\end{aligned} \tag{4.20}$$

The two error components are due to rank deficiency and noise, respectively. (Note that if A has full column rank, then $A^+A = I$ and $\|\mathbf{e}_r\|^2 = 0$; similarly, if there is no noise, then $\mathbf{n} = \mathbf{0}$ and $\|\mathbf{e}_n\|^2 = 0$.)

First, consider the error due to rank deficiency. Using SVD, let $A = USV^H$, giving $A^+A = VS^+U^HUSV^H = VS^+SV^H = V\Lambda V^H$, where Λ is a diagonal matrix with ones and zeros

on its diagonal corresponding to the rank of A; that is, Λ specifies the columns of V (i.e. the right singular vectors) which are spanned by the rowspace of A. Now,

$$\begin{aligned}\|\mathbf{e}_r\|^2 &= \mathbf{f}^H(VV^H - V\Lambda V^H)\mathbf{f} \\ &= (V^H\mathbf{f})^H(I-\Lambda)(V^H\mathbf{f}).\end{aligned} \tag{4.21}$$

Note that $V^H\mathbf{f}$ is simply a projection of $\mathbf{f}$ onto the columns of V. Therefore, the error power due to rank deficiency is given by the power of $\mathbf{f}$ that is present in the nullspace of A. This is satisfying intuitively: the nullspace component of $\mathbf{f}$ is orthogonal to the rowspace of A, and is therefore lost by the (direct) imaging process. The effect of truncation in the reconstruction is to set the elements of S^+ corresponding to the smallest singular values to zero. This is reflected in Λ and results in a higher error due to rank deficiency, even if A has full column rank.

Now consider the error due to noise. Since $(AA^H)^+$ is Hermitian, it is unitarily diagonalisable [Str88] and may be written as $V\Lambda V^H$ where V is a unitary matrix and Λ is a diagonal matrix of the eigenvalues of $(AA^H)^+$. If $\mathbf{n}$ is zero-mean, white, Gaussian noise with variance s^2, then

$$\begin{aligned}E\left\{\|\mathbf{e}_n\|^2\right\} &= E\left\{\mathbf{n}^H V\Lambda V^H\mathbf{n}\right\} \\ &= E\left\{(V^H\mathbf{n})^H\Lambda(V^H\mathbf{n})\right\}.\end{aligned}$$

Since V is unitary, the vector $(V^H\mathbf{n})$ has the same statistics as $\mathbf{n}$. Let the elements of this new vector be $u_1, u_2, \ldots, u_p$, and let the elements of the diagonal matrix Λ be $\lambda_1, \lambda_2, \ldots, \lambda_p$. Then,

$$\begin{aligned}E\left\{\|\mathbf{e}_n\|^2\right\} &= E\left\{\bar{u}_1\lambda_1 u_1 + \bar{u}_2\lambda_2 u_2 + \cdots + \bar{u}_p\lambda_p u_p\right\} \\ &= E\left\{\lambda_1|u_1|^2 + \lambda_2|u_2|^2 + \cdots + \lambda_p|u_p|^2\right\} \\ &= \lambda_1 E\left\{|u_1|^2\right\} + \lambda_2 E\left\{|u_2|^2\right\} + \cdots + \lambda_p E\left\{|u_p|^2\right\}.\end{aligned}$$

The expected value of each element $|u_i|^2$ is simply the variance, s^2, and the sum of all eigenvalues is simply the trace, giving

$$\begin{aligned}E\left\{\|\mathbf{e}_n\|^2\right\} &= s^2\sum_{i=1}^{p}\lambda_i \\ &= s^2\,\mathrm{trace}(AA^H)^+ \\ &= s^2\sum_{i=1}^{\min(p,q)}\frac{1}{\sigma_i^2},\end{aligned} \tag{4.22}$$

where $\sigma_1, \sigma_2, \ldots, \sigma_{\min(p,q)}$ are the singular values of A (see Section 2.3.2).

While Eq. (4.21) depends on the true object, $\mathbf{f}$, Eq. (4.22) depends only on matrix A, making it more desirable as a metric. Although the total error power comprises two components, it is usual to consider the situation where perfect reconstruction would be possible in the absence of noise (i.e. A has full column rank), in which case $\|\mathbf{e}_r\|^2$ is zero and the only source of error is the noise component, $\|\mathbf{e}_n\|^2$. (Note that if truncation of the reconstruction is performed (see Section 2.6.1), the error due to rank deficiency will be increased while the error due to noise will be reduced, generally giving a net decrease in total error.) With these justifications for concentrating on the noise component of the error, the *trace metric* is defined as

$$\text{tmetric}(A) = \sum_{i=0}^{\min(p,q)} \frac{1}{\sigma_i^2} = \begin{cases} \text{trace}(A^H A)^{-1} & p \geq q \\ \text{trace}(AA^H)^{-1} & p < q. \end{cases} \tag{4.23}$$

This is equivalent to the criterion used by Reeves *et al.* mentioned above [RH95, GR00]. It is an excellent measure of the conditioning of the reconstruction, partly because in contrast to $\text{cond}(A)$, $\text{tmetric}(A)$ is a function of *all* singular values. Note that $W_N[\alpha|\beta]$ has full row or column rank (for $p \leq q$ or $p \geq q$, respectively) if and only if the trace metric is finite.

Comparisons between problem sizes

As well as comparing performance among different sampling sets of the same size, it is desirable to make comparisons as the problem size varies (i.e. changes in p, q, and/or N). First, consider the following result:

Theorem 4.1. *Let* $A \in \mathbb{C}^{p\times q}$, $\mathbf{u} \in \mathbb{C}^p$, *and* $\mathbf{v} \in \mathbb{C}^q$. *Let* $B = \left[\begin{array}{c} A \\ \hdashline \mathbf{v}^H \end{array}\right]$ *and* $C = [\,A \,\vdots\, \mathbf{u}\,]$.

(a) If $p \geq q$ *then* $\text{tmetric}(B) \leq \text{tmetric}(A)$, *and if* $p < q$ *then* $\text{tmetric}(B) \geq \text{tmetric}(A)$.

(b) If $p \leq q$ *then* $\text{tmetric}(C) \leq \text{tmetric}(A)$, *and if* $p > q$ *then* $\text{tmetric}(C) \geq \text{tmetric}(A)$.

Proof. (a) If $p \geq q$, then let A and B have singular values $a_1 \geq a_2 \geq \cdots \geq a_q$ and $b_1 \geq b_2 \geq \cdots \geq b_q$, respectively. A is a submatrix of B, so by Thompson's theorem of interlacing singular values [Tho72],

$$\begin{aligned} b_i &\geq a_i \quad i = 1, 2, \ldots, q \\ a_i &\geq b_{i+1} \quad i = 1, 2, \ldots, q-1. \end{aligned} \tag{4.24}$$

Since $\text{tmetric}(A) = \sum_{i=1}^{q} 1/a_i^2$ and $\text{tmetric}(B) = \sum_{i=1}^{q} 1/b_i^2$, it can be deduced from the first part of Eq. (4.24) that

$$\begin{aligned} \frac{1}{b_i^2} &\leq \frac{1}{a_i^2} \quad i = 1, 2, \ldots, q \\ \Rightarrow \text{tmetric}(B) = \sum_{i=1}^{q} \frac{1}{b_i^2} &\leq \sum_{i=1}^{q} \frac{1}{a_i^2} = \text{tmetric}(A) \end{aligned} \tag{4.25}$$

If $p < q$ then let A and B have singular values $a_1 \geq a_2 \geq \cdots \geq a_p$ and $b_1 \geq b_2 \geq \cdots \geq b_{p+1}$, respectively. Thompson's theorem gives

$$\begin{aligned} b_i &\geq a_i \quad i = 1, 2, \ldots, p \\ a_i &\geq b_{i+1} \quad i = 1, 2, \ldots, p. \end{aligned} \tag{4.26}$$

In this case, $\text{tmetric}(A) = \sum_{i=1}^{p} 1/a_i^2$ and $\text{tmetric}(B) = \sum_{i=1}^{p+1} 1/b_i^2$, and so from the second part of Eq. (4.26),

$$\begin{aligned} \frac{1}{b_{i+1}^2} &\geq \frac{1}{a_i^2} \quad i = 1, 2, \ldots, p \\ \Rightarrow \sum_{i=2}^{p+1} \frac{1}{b_i^2} &\geq \sum_{i=1}^{p} \frac{1}{a_i^2} \\ \Rightarrow \text{tmetric}(B) = \frac{1}{b_1^2} + \sum_{i=2}^{p+1} \frac{1}{b_i^2} &\geq \sum_{i=1}^{p} \frac{1}{a_i^2} = \text{tmetric}(A) \end{aligned} \tag{4.27}$$

(b) This follows analogously to (a) with p and q swapped, and C and c_i substituted for B and b_i, respectively. □

To make fair comparisons between different problem sizes, consider the use of the expected value of the relative mean square error per image point as a benchmark:

$$\frac{E\left\{\|\mathbf{e}\|^2\right\}/q}{\|\mathbf{f}\|^2/q}. \tag{4.28}$$

If an extra sample is measured ($p \uparrow$), any change in Eq. (4.28) is only due to a change in $E\left\{\|\mathbf{e}\|^2\right\}$. It is therefore fair to compare problems with different numbers of measurements using the trace metric. Note that the rank deficiency component of the error may also change, but this investigation is concerned with the sensitivity to noise as reflected in $E\left\{\|\mathbf{e}_n\|^2\right\}$. Intuitively, adding more samples to a fully or over-determined system should not be able to increase the error, and this is confirmed by part (a) of Theorem 4.1 above when $p \geq q$. If $p < q$, the error due to noise actually increases (!), but this occurs concurrently with a decrease in the error due to rank deficiency, probably decreasing the total error.

If the size of the support is increased ($q \uparrow$), then the signal power is unaffected, and the change in q affects the numerator and the denominator identically. Any change in Eq. (4.28) is hence due to $E\left\{\|\mathbf{e}\|^2\right\}$, meaning it is fair to compare problems with different numbers of unknowns using the trace metric. This time, intuition suggests that the error power should

be unable to decrease since there are now more unknowns, and this is reflected in part (b) of Theorem 4.1 when $p > q$.

When the size of the problem (N) is changed, the elements of A are affected since A describes the relationship between the object and image. Specifically, the elements of $W_{\mathrm{N}}[\alpha|\beta]$, as shown in Eq. (4.7), are affected in two ways: the argument of the exponential changes due to the new relative positions of sample locations, and the $1/\sqrt{N}$ normalising factor is altered. The latter change should be reflected in f so that the matrix equation relationship still holds; namely, a change from N to N′ means that f must be multiplied by N'/N, where $N' = N'_0 N'_1 \cdots N'_{M-1}$. Therefore, the trace metrics for the two problem sizes must be divided by N and N', respectively, for a fair comparison.

The following example demonstrates a comparison of conditioning between systems as p is varied while also illustrating the benefit of using $\mathrm{tmetric}(A)$ instead $\mathrm{cond}(A)$.

Example 4.1. Let $M = 1$, $N = 8$, $\beta = \{0, 1, 2, 3\}$, $\alpha = \{0, 2, 4, 6\}$, and $\alpha' = \{0, 1, 2, 4, 6\}$. First consider the system that has support set β and sampling set α, giving $p = q = 4$. The singular values of $A = W_N[\alpha|\beta]$ are

$$\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}},$$

which gives $\mathrm{cond}(A) = 1$ and $\mathrm{tmetric}(A) = 8$. Now consider the system with support set β and sampling set α', giving $p' = 5$ and $q' = 4$. Note that the only difference between this and the first system is one extra element in the sampling set. The singular values of $A' = W_N[\alpha'|\beta]$ are

$$1, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}},$$

which gives $\mathrm{cond}(A') = 1.41$ and $\mathrm{tmetric}(A') = 7$.

An intuitive interpretation of conditioning suggests that conditioning should reflect performance under noise: if one system generally gives lower error due to noise than an alternative system, it is better conditioned. However, the second system is worse conditioned than the first according to the condition number even though it has had one extra sample added! This is due to the fact that the condition number reflects a bound on the noise amplification and is therefore necessarily conservative. It is fair to compare trace metrics between these systems since only p has been altered, and the extra sample added has resulted in a slightly better conditioned system. This is due to the fact that the trace metric involves a statistical average (the expected value) of the noise power. It is therefore argued that the trace metric gives a better intuitive interpretation of conditioning for the type of system analysed here.

MATLAB® was used to create the matrices A and A' and an arbitrary vector, $\mathbf{x} \in \mathbb{C}^q$. Measurement vectors $\mathbf{b} = A\mathbf{x}$ and $\mathbf{b}' = A\mathbf{x}'$ were then computed. Zero-mean, unit-variance, complex, white, Gaussian noise was added to the measurement vectors and the least-squares solutions computed. Averaged over 10,000 noise realisations, the norm-squared values of the error vectors were 7.999 and 7.033, confirming the numerical meaning underlying the trace metric. ▲

The need to divide the trace metric by N for comparisons between problem sizes could be avoided by incorporating this factor into the definition. In addition, note that the sampling situation in which no samples are omitted and no prior support information is assumed is characterised by the complete Fourier matrix, which has trace metric N. Once normalised by N, therefore, the trace metric would have a simple interpretation as the noise penalty (expressed in decibels for example) associated with sampling below the Nyquist density and using prior knowledge of the support for reconstruction. The standard trace metric is predominantly used in this thesis, since most results involve comparing trace metrics resulting from different sampling schemes and it is therefore their *relative* magnitudes that are important.

4.4.2 Relationship between 1-D and M-D

If the individual elements of the size of a multi-dimensional problem are pairwise relatively prime, the M-D problem can be represented by a 1-D problem. The ability to focus on the Fourier matrix for 1-D sequences is clearly an advantage in simplifying analysis of the problem.

Consider the sets $\alpha \in Q_{p,N}$ and $\boldsymbol{\alpha}' \in Q_{p,\mathbf{N}}$, which represent locations in a 1-D and an M-D sequence, respectively, where $N = N_0 N_1 \cdots N_{M-1}$ and $(N_0, N_1, \ldots, N_{M-1}) = 1$. In a similar fashion to Bresler and Feng [BF96], define a method for unique conversion between α and $\boldsymbol{\alpha}'$: $\boldsymbol{\alpha}' = \alpha \mathbf{1} \bmod \mathbf{N}$. The pairwise relatively prime restriction ensures this mapping of elements is one-to-one, and the reverse conversion (from M-D to 1-D) is easily computed via the Chinese remainder theorem (Theorem 2.4).

Theorem 4.2. *Let $\alpha \in Q_{p,N}$ and $\beta \in Q_{q,N}$, and let $\mathbf{N}$ represent the size of M-D space such that $N = N_0 N_1 \cdots N_{M-1}$, with $(N_0, N_1, \ldots, N_{M-1}) = 1$. If*

$$\boldsymbol{\alpha}' = \alpha \mathbf{1} \bmod \mathbf{N}, \tag{4.29}$$

$$\boldsymbol{\beta}' = \beta \mathbf{1} \bmod \mathbf{N}, \text{ and} \tag{4.30}$$

$$\hat{\beta} = \beta \left[N_1 N_2 \cdots N_{M-1} + N_0 N_2 \cdots N_{M-1} + \cdots + N_0 N_1 \cdots N_{M-2} \right] \bmod N, \tag{4.31}$$

then $W_N[\alpha|\hat{\beta}]$ and $W_{\mathbf{N}}[\boldsymbol{\alpha}'|\boldsymbol{\beta}']$ have identical singular values.

Proof. Let $A = W_N[\alpha|\hat{\beta}]$ and $B = W_{\mathbf{N}}[\alpha'|\beta']$ be submatrices of the Fourier matrices for 1-D and M-D sequences, respectively. Using Eq. (4.31) to expand $\hat{\beta}_j$,

$$\begin{aligned}(A)_{kl} &= \exp\left(\frac{-j2\pi\alpha_k\hat{\beta}_l}{N}\right) \\ &= \exp\left(\frac{-j2\pi\alpha_k\beta_l}{N}\left[N_1N_2\cdots N_{M-1} + N_0N_2\cdots N_{M-1} + \cdots + N_0N_1\cdots N_{M-2}\right]\right).\end{aligned} \tag{4.32}$$

Using Eqs. (4.29) and (4.30),

$$(B)_{kl} = \exp\left(-j2\pi\left[\frac{(\alpha_k \bmod N_0)(\beta_l \bmod N_0)}{N_0} + \cdots + \frac{(\alpha_k \bmod N_{M-1})(\beta_l \bmod N_{M-1})}{N_{M-1}}\right]\right),$$

and by Theorem 2.3,

$$= \exp\left(-j2\pi\left[\frac{\alpha_k\beta_l}{N_0} + \cdots + \frac{\alpha_k\beta_l}{N_{M-1}}\right]\right)$$

$$= \exp\left(\frac{-j2\pi\alpha_k\beta_l}{N}\left[N_1N_2\cdots N_{M-1} + N_0N_2\cdots N_{M-1} + \cdots + N_0N_1\cdots N_{M-2}\right]\right) = (A)_{kl}. \tag{4.33}$$

Therefore, A and B have identical singular values. The pairwise relatively prime restriction is not required for this result to hold, but to ensure a one-to-one mapping of 1-D sample locations onto M-D sample locations, $N_1N_2\cdots N_{M-1}+N_0N_2\cdots N_{M-1}+\cdots+N_0N_1\cdots N_{M-2}$ and N must be pairwise relatively prime, or equivalently $N_0, N_1, \ldots, N_{M-1}$ must be pairwise relatively prime. □

This is illustrated by the following example.

Example 4.2. Let $N = 15$, $M = 2$, $\mathbf{N} = \binom{5}{3}$, $\alpha = \{0, 2, 7, 8\}$, and $\beta = \{0, 4, 7\}$. From Eqs. (4.29) and (4.30),

$$\boldsymbol{\alpha}' = \left\{\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}2\\2\end{pmatrix}, \begin{pmatrix}2\\1\end{pmatrix}, \begin{pmatrix}3\\2\end{pmatrix}\right\}, \text{ and}$$

$$\boldsymbol{\beta}' = \left\{\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}4\\1\end{pmatrix}, \begin{pmatrix}2\\1\end{pmatrix}\right\},$$

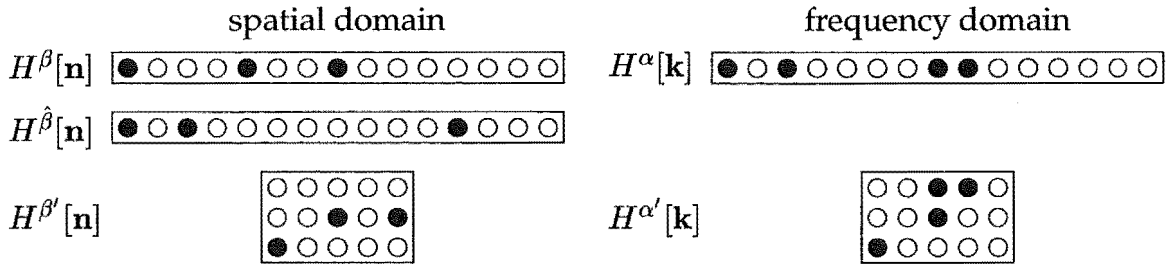


Figure 4.2 Sample sets involved in Example 4.2.

and from Eq. (4.31),

$$\hat{\beta} = \{0, 2, 11\}\,.$$

The sample sets involved are illustrated in Fig. 4.2; note that $p = 4$ and $q = 3$. The 1-D problem with α and $\hat{\beta}$ now has the same singular values as the 2-D problem with α' and β', and hence the same trace metric. ▲

In other words, provided that the elements of the size of M-D space are pairwise relatively prime, the M-D problem may be analysed by considering its equivalent 1-D version.

4.4.3 Equivalent problems

In general, the number of possible sampling sets $\boldsymbol{\alpha} \in Q_{p,\mathbf{N}}$ and support sets $\boldsymbol{\beta} \in Q_{q,\mathbf{N}}$ are given by Theorem 2.6 as $\binom{N}{p}$ and $\binom{N}{q}$, respectively. This means there are $\binom{N}{p} \times \binom{N}{q}$ possibilities for $W_{\mathbf{N}}[\alpha|\beta]$! Two theorems are developed in this section that show equivalences between different problems, simplifying analyses and facilitating practical algorithms.

Definition 4.4. Two sets $\boldsymbol{\alpha}, \boldsymbol{\alpha}' \in Q_{p,\mathbf{N}}$ are related by a *cyclic shift* if $\boldsymbol{\alpha}' = (\boldsymbol{\alpha} + \mathbf{a}) \bmod \mathbf{N}$ for some $\mathbf{a} \in \mathbb{Z}^M$.

Definition 4.5. Two sets $\boldsymbol{\alpha}, \boldsymbol{\alpha}' \in Q_{p,\mathbf{N}}$ are related by a *mirror image* if $\boldsymbol{\alpha}' = (\mathbf{N} - \boldsymbol{\alpha}) \bmod \mathbf{N}$.

Theorem 4.3. *Let $\boldsymbol{\alpha} \in Q_{p,\mathbf{N}}$ and $\boldsymbol{\beta} \in Q_{q,\mathbf{N}}$. A cyclic shift and/or mirror image of $\boldsymbol{\alpha}$ and/or $\boldsymbol{\beta}$ does not affect the singular values of $W_{\mathbf{N}}[\alpha|\beta]$.*

Proof. (a) Cyclic shift of $\boldsymbol{\beta}$: Let $\boldsymbol{\beta}' = (\boldsymbol{\beta} + \mathbf{b}) \bmod \mathbf{N}$ for some $\mathbf{b} \in \mathbb{Z}^M$, $A = W_{\mathbf{N}}[\alpha|\boldsymbol{\beta}]$ and $B = W_{\mathbf{N}}[\alpha|\boldsymbol{\beta}']$, then

$$(A^H A)_{kl} = \frac{1}{N} \sum_{i=0}^{p-1} \prod_{j=0}^{M-1} w_{N_j}^{(\boldsymbol{\alpha}_i)_j[(\boldsymbol{\beta}_k)_j - (\boldsymbol{\beta}_l)_j]}, \tag{4.34}$$

$$\begin{aligned}(B^H B)_{kl} &= \frac{1}{N} \sum_{i=0}^{p-1} \prod_{j=0}^{M-1} w_{N_j}^{(\boldsymbol{\alpha}_i)_j[(\boldsymbol{\beta}_k + \mathbf{b})_j - (\boldsymbol{\beta}_l + \mathbf{b})_j]} \\ &= \frac{1}{N} \sum_{i=0}^{p-1} \prod_{j=0}^{M-1} w_{N_j}^{(\boldsymbol{\alpha}_i)_j[(\boldsymbol{\beta}_k)_j + (\mathbf{b})_j - (\boldsymbol{\beta}_l)_j - (\mathbf{b})_j]} \\ &= \frac{1}{N} \sum_{i=0}^{p-1} \prod_{j=0}^{M-1} w_{N_j}^{(\boldsymbol{\alpha}_i)_j[(\boldsymbol{\beta}_k)_j - (\boldsymbol{\beta}_l)_j]} = (A^H A)_{kl}.\end{aligned} \tag{4.35}$$

$A^H A$ and $B^H B$ therefore have identical eigenvalues and hence A and B have identical singular values.

(b) Mirror image of $\boldsymbol{\beta}$: Let $\boldsymbol{\beta}' = (\mathbf{N} - \boldsymbol{\beta}) \bmod \mathbf{N}$, so that

$$\begin{aligned}(B^H B)_{kl} &= \frac{1}{N} \sum_{i=0}^{p-1} \prod_{j=0}^{M-1} w_{N_j}^{(\boldsymbol{\alpha}_i)_j[(\mathbf{N} - \boldsymbol{\beta}_k)_j - (\mathbf{N} - \boldsymbol{\beta}_l)_j]} \\ &= \frac{1}{N} \sum_{i=0}^{p-1} \prod_{j=0}^{M-1} w_{N_j}^{(\boldsymbol{\alpha}_i)_j[(\mathbf{N})_j - (\boldsymbol{\beta}_k)_j - (\mathbf{N})_j + (\boldsymbol{\beta}_l)_j]} \\ &= \frac{1}{N} \sum_{i=0}^{p-1} \prod_{j=0}^{M-1} w_{N_j}^{(\boldsymbol{\alpha}_i)_j[(\boldsymbol{\beta}_l)_j - (\boldsymbol{\beta}_k)_j]} = \overline{(A^H A)_{kl}}.\end{aligned} \tag{4.36}$$

The eigenvalues of a conjugated matrix are the conjugates of the eigenvalues, but since $A^H A$ and $B^H B$ are Hermitian, their eigenvalues are real and therefore identical, and hence the singular values of A and B are also identical. The proof for α is identical. □

This result was noted in Ref. [FB96] and proved for cyclic shifts in Ref. [GR00]. The ability to disregard cyclic shifts and mirror images is a valuable aid in analysing and solving the problem. The number of possible sets $\alpha \in Q_{p,\mathbf{N}}$ that need be considered is therefore reduced from $\binom{N}{p}$ and may be calculated exactly using Theorem 2.7. Let $\hat{Q}_{p,\mathbf{N}} \subseteq Q_{p,\mathbf{N}}$ contain only the 'unique' members of $Q_{p,\mathbf{N}}$; that is, all members of $Q_{p,\mathbf{N}}$ are related by a cyclic shift and/or mirror image to a member of $\hat{Q}_{p,\mathbf{N}}$ but no two members of $\hat{Q}_{p,\mathbf{N}}$ are related in this way.

Example 4.3. Let $M = 1$. The number of unique members of $Q_{p,N}$, namely $\left|\hat{Q}_{p,N}\right|$, is calculated here using Theorem 2.7. The N sampling locations $0, 1, \ldots, N-1$ are to be 'painted' with two colours, 'measured' and 'not measured', which are assigned weights x

and y, respectively. A group, G, contains cyclic shift and mirror image permutations of the N sampling locations.

For cyclic shifts, there are N possible permutations, relating to cyclic shifts of $a \in \{0, 1, \ldots, N-1\}$. A cyclic shift of zero is simply the identity, and has N cycles:

$$\begin{pmatrix} 0 & 1 & \ldots & N-1 \\ 0 & 1 & \ldots & N-1 \end{pmatrix} = (0)(1)\cdots(N-1). \tag{4.37}$$

Generally, if e is the starting element in the cycle, the number of terms in the cycle, n, is given by the smallest solution to $e + na \equiv e \pmod{N}$ or equivalently $na \equiv 0 \pmod{N}$, such that $n \in \{1, 2, \ldots, N\}$. If $(a, N) = d$, then $n = N/d$, and so there are exactly d cycles of N/d terms. For example, $N = 10$, $a = 4$:

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 4 & 5 & 6 & 7 & 8 & 9 & 0 & 1 & 2 & 3 \end{pmatrix} = (0\,4\,8\,2\,6)(1\,5\,9\,3\,7). \tag{4.38}$$

For mirror images, there are again N possible permutations, relating to a mirror image together with a cyclic shift of $a \in \{0, 1, \ldots, N-1\}$. The number of cycles depends on whether N is odd or even. If N is odd, then there is one cycle of 1 and $(N-1)/2$ cycles of 2, regardless of a. For example, $N = 9$, $a = 2$:

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 2 & 1 & 0 & 8 & 7 & 6 & 5 & 4 & 3 \end{pmatrix} = (0\,2)(1)(3\,8)(4\,7)(5\,6). \tag{4.39}$$

On the other hand, if N is even, then the result depends on whether a is even or odd: if a is even, there are 2 cycles of 1 and $(N-2)/2$ cycles of 2, and if a is odd, there are $N/2$ cycles of 2 and none of 1. For example, $N = 8$, $a = 3$ and $a = 4$:

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 2 & 1 & 0 & 7 & 6 & 5 & 4 \end{pmatrix} = (0\,3)(1\,2)(4\,7)(5\,6), \tag{4.40}$$

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 4 & 3 & 2 & 1 & 0 & 7 & 6 & 5 \end{pmatrix} = (0\,4)(1\,3)(2)(5\,7)(6). \tag{4.41}$$

Therefore, the pattern inventory for N odd is

$$\frac{1}{2N}\left\{(x+y)^N + \left[\sum_{a=1}^{N-1}(x^{\frac{N}{(a)N}} + y^{\frac{N}{(a)N}})^{(a)N}\right] + N(x+y)(x^2+y^2)^{\frac{N-1}{2}}\right\}, \tag{4.42}$$

and for N even is

$$\frac{1}{2N}\left\{(x+y)^N+\left[\sum_{a=1}^{N-1}(x^{\frac{N}{(a)N}}+y^{\frac{N}{(a)N}})^{(a)N}\right]+\frac{N}{2}(x+y)^2(x^2+y^2)^{\frac{N-2}{2}}+\frac{N}{2}(x^2+y^2)^{\frac{N}{2}}\right\}. \quad (4.43)$$

The number of unique sets $\alpha \in Q_{p,N}$ is simply the coefficient of the term $x^p y^{N-p}$ in the appropriate version of the above equation. Table 4.1 lists the results for $2 \leq N \leq 22$ and $0 \leq p \leq N$. ▲

The previous theorem reduced the number of sets $\alpha \in Q_{p,\mathbf{N}}$ and $\beta \in Q_{q,\mathbf{N}}$ that need be considered at all. The following theorem shows an equivalence between two *pairs* of sampling and support sets. While this does not reduce the overall number of sets, it does reduce the number of combinations of $W_{\mathbf{N}}[\alpha|\beta]$ that have distinct properties. As discussed in relation to sample-selection criteria in Section 6.6.1, this has implications for the attributes of a combination of sampling and support sets that are relevant for conditioning.

Theorem 4.4. *Let* $\alpha \in Q_{p,\mathbf{N}}$ *and* $\beta \in Q_{q,\mathbf{N}}$, *and let* $\alpha' = (\mathbf{a} \circ \alpha) \bmod \mathbf{N}$ *and* $\beta' = (\mathbf{b} \circ \beta) \bmod \mathbf{N}$, *where* $\mathbf{a}, \mathbf{b} \in \mathbb{Z}^M$ *and* $a_j b_j \equiv \pm 1 \pmod{N_j}$, $j = 0, 1, \ldots, M-1$. *Then,* $W_{\mathbf{N}}[\alpha|\beta]$ *and* $W_{\mathbf{N}}[\alpha'|\beta']$ *have identical singular values.*

Proof. Let $A = W_{\mathbf{N}}[\alpha|\beta]$ and $B = W_{\mathbf{N}}[\alpha'|\beta']$, so that

$$(A^H A)_{kl} = \frac{1}{N}\sum_{i=0}^{p-1}\prod_{j=0}^{M-1} w_{N_j}^{(\alpha_i)_j[(\beta_k)_j-(\beta_l)_j]}, \quad (4.44)$$

$$(B^H B)_{kl} = \frac{1}{N}\sum_{i=0}^{p-1}\prod_{j=0}^{M-1} w_{N_j}^{a_j(\alpha_i)_j[b_j(\beta_k)_j-b_j(\beta_l)_j]}$$

$$= \frac{1}{N}\sum_{i=0}^{p-1}\prod_{j=0}^{M-1} w_{N_j}^{a_j b_j(\alpha_i)_j[(\beta_k)_j-(\beta_l)_j]}$$

and by Theorem 2.3,

$$= \frac{1}{N}\sum_{i=0}^{p-1}\prod_{j=0}^{M-1} w_{N_j}^{\pm(\alpha_i)_j[(\beta_k)_j-(\beta_l)_j]} = (A^H A)_{kl} \text{ or } \overline{(A^H A)_{kl}}. \quad (4.45)$$

Hence $A^H A$ and $B^H B$ have identical eigenvalues, and A and B have identical singular values. □

Note that in Theorem 4.3, *either* α or β (or both) can be modified, but in Theorem 4.4 *both* must be modified according to the given relationship.

Table 4.1 Number of unique[a] sets $\alpha \in Q_{p,N}$

N	p	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
2		1	1	1																				
3		1	1	1	1																			
4		1	1	2	1	1																		
5		1	1	2	2	1	1																	
6		1	1	3	3	3	1	1																
7		1	1	3	4	4	3	1	1															
8		1	1	4	5	8	5	4	1	1														
9		1	1	4	7	10	10	7	4	1	1													
10		1	1	5	8	16	16	16	8	5	1	1												
11		1	1	5	10	20	26	26	20	10	5	1	1											
12		1	1	6	12	29	38	50	38	29	12	6	1	1										
13		1	1	6	14	35	57	76	76	57	35	14	6	1	1									
14		1	1	7	16	47	79	126	133	126	79	47	16	7	1	1								
15		1	1	7	19	56	111	185	232	232	185	111	56	19	7	1	1							
16		1	1	8	21	72	147	280	375	440	375	280	147	72	21	8	1	1						
17		1	1	8	24	84	196	392	600	750	750	600	392	196	84	24	8	1	1					
18		1	1	9	27	104	252	561	912	1282	1387	1282	912	561	252	104	27	9	1	1				
19		1	1	9	30	120	324	756	1368	2052	2494	2494	2052	1368	756	324	120	30	9	1	1			
20		1	1	10	33	145	406	1032	1980	3260	4262	4752	4262	3260	1980	1032	406	145	33	10	1	1		
21		1	1	10	37	165	507	1353	2829	4950	7105	8524	8524	7105	4950	2829	1353	507	165	37	10	1	1	
22		1	1	11	40	195	621	1782	3936	7440	11410	14938	16159	14938	11410	7440	3936	1782	621	195	40	11	1	1

[a] 'unique' sets are those not related by a cyclic shift and/or mirror image

4.4.4 Singular values

Theorems are presented here that give properties of the singular values of $W_{\mathbf{N}}[\alpha|\beta]$.

Theorem 4.5. *Let $\alpha \in Q_{p,\mathbf{N}}$ and let $\beta \in Q_{q,\mathbf{N}}$. With the singular values of $W_{\mathbf{N}}[\alpha|\beta]$ designated $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min(p,q)}$, then*

$$\begin{aligned} \sigma_i &= 1 \qquad i \leq p+q-N \\ \sigma_i &< 1 \qquad i > p+q-N. \end{aligned} \tag{4.46}$$

Proof. The singular values of $W_{\mathbf{N}}$ are given by the positive square roots of the eigenvalues of $W_{\mathbf{N}}^H W_{\mathbf{N}} = I$; that is, $W_{\mathbf{N}}$ has N singular values of one. By Thompson's theorem of interlacing singular values [Tho72],

$$\begin{aligned} 1 &\geq \sigma_i \quad i = 1, 2, \ldots, \min(p,q) \\ \sigma_i &\geq 1 \quad i = 1, 2, \ldots, p+q-N. \end{aligned} \tag{4.47}$$

Together, these two conditions produce the required result. □

Theorem 4.6. *Let $\alpha \in Q_{p,\mathbf{N}}$ and let $\beta \in Q_{q,\mathbf{N}}$. With the singular values of $W_{\mathbf{N}}[\alpha|\beta]$ designated $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min(p,q)}$,*

$$\sum_{i=1}^{\min(p,q)} \sigma_i^2 = \frac{pq}{N}. \tag{4.48}$$

Proof. Let $A = W_{\mathbf{N}}[\alpha|\beta]$. Then,

$$\begin{aligned} \operatorname{trace}(A^H A) &= \sum_{k=0}^{q-1} \frac{1}{N} \sum_{i=0}^{p-1} \prod_{j=0}^{M-1} w_{N_j}^{(\alpha_i)_j[(\beta_k)_j - (\beta_k)_j]}, \\ &= \frac{1}{N} \sum_{k=0}^{q-1} \sum_{i=0}^{p-1} \prod_{j=0}^{M-1} w_{N_j}^{0}, \\ &= \frac{1}{N} \sum_{k=0}^{q-1} \sum_{i=0}^{p-1} 1, \\ &= \frac{pq}{N} \end{aligned} \tag{4.49}$$

The trace of a square matrix is equal to the sum of its eigenvalues [Str88], and the eigenvalues of $A^H A$ are the squares of the singular values of A, so the sum of the eigenvalues equals the sum of the squares of the singular values. □

Theorem 4.7. *Let $\alpha \in Q_{p,\mathbf{N}}$, $\beta \in Q_{q,\mathbf{N}}$, and $\gamma = N - p - q \geq 0$. Then, $W_{\mathbf{N}}[\alpha|\beta]$ has singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min(p,q)}$ if and only if $W_{\mathbf{N}}[\bar{\alpha}|\bar{\beta}]$ has singular values $\underbrace{1 = 1 = \cdots = 1}_{\gamma} \geq \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min(p,q)}$.*

Proof. Consider the following reordering and partitioning of the Fourier matrix:

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} W_{\mathbf{N}}[\alpha|\beta] & W_{\mathbf{N}}[\alpha|\bar{\beta}] \\ W_{\mathbf{N}}[\bar{\alpha}|\beta] & W_{\mathbf{N}}[\bar{\alpha}|\bar{\beta}] \end{pmatrix}. \tag{4.50}$$

Since $AA^H = A^H A = I$, $A_{11}A_{11}^H + A_{12}A_{12}^H = I \in \mathbb{Z}^{p\times p}$ and $A_{12}^H A_{12} + A_{22}^H A_{22} = I \in \mathbb{Z}^{(N-q)\times(N-q)}$. For any $m \times n$ matrix A, the eigenvalues of AA^H and $A^H A$ are equal, except that the larger of the two matrices has an additional $|m - n|$ zero eigenvalues. Assume that $p \leq q$ and therefore consider the matrices $A_{11}A_{11}^H$ and $A_{22}^H A_{22}$, which have p and $(N - q)$ eigenvalues, respectively, the same numbers of singular values as A_{11} and A_{22} have, respectively. Note that the assumption $N - p - q \geq 0$ ensures that A_{22} has at least as many singular values as A_{11}. For the case $p > q$, consider $A_{11}^H A_{11}$ and $A_{22}A_{22}^H$ instead.

Let $\lambda_1, \ldots, \lambda_p$ denote the eigenvalues of $A_{11}A_{11}^H$. Then the eigenvalues of $A_{12}A_{12}^H = I - A_{11}A_{11}^H$ are clearly $\lambda_k' = 1 - \lambda_k$ for $k = 1, \ldots, p$. The eigenvalues of $A_{12}^H A_{12}$ are therefore $\lambda_k'' = \lambda_k'$ for $k = 1, \ldots, p$ and $\lambda_k'' = 0$ for $k = p + 1, \ldots, N - q$. Finally, the eigenvalues of $A_{22}^H A_{22} = I - A_{12}^H A_{12}$ are $\lambda_k''' = 1 - \lambda_k''$ for $k = 1, \ldots, N - q$, that is, $\lambda_k''' = \lambda_k$ for $k = 1, \ldots, p$ and $\lambda_k''' = 1$ for $k = p + 1, \ldots, N - q$. The singular values of A_{22} are therefore equal to the singular values of A_{11}, with an additional $N - q - p$ of value one. □

As a consequence of Theorem 4.7, any sampling problem may be approached by considering its complement, since their singular values are identical (except for a number with value unity). In particular, consider the following corollary:

Corollary. *Let $\beta \in Q_{q,\mathbf{N}}$. If*

$$\alpha = \operatorname*{argmin}_{\alpha' \in Q_{p,\mathbf{N}}} \operatorname{tmetric}(W_N[\alpha'|\beta]), \tag{4.51}$$

then

$$\bar{\alpha} = \operatorname*{argmin}_{\alpha' \in Q_{N-p,\mathbf{N}}} \operatorname{tmetric}(W_N[\alpha'|\bar{\beta}]). \tag{4.52}$$

4.4.5 The case N prime

The discrete nature of the problem has interesting implications concerning the factors of N and its relationship with p, namely (p, N). The importance of factors is investigated in Section 6.3.3, but note here the intriguing result that occurs in the complete absence of factors of N:

Theorem 4.8. *If N is prime, then any $p \times p$ submatrix of the Fourier matrix for 1-D sequences, W_N, is non-singular.*

See Ref. [AIGBD94] for proof. It should be stressed that although the matrix is non-singular, it may be so badly conditioned to render it effectively singular. Nevertheless, this result certainly suggests that factors play a vital role in determining singularity.

4.4.6 Contiguous support

A contiguous support arises in traditional superresolution problems such as that presented in Ref. [WN94], and Bresler *et al.* investigated a contiguous (or "bunched") pattern in the context of universal patterns [BF96] (introduced in the next section). For these reasons, it is an interesting set to study. The following definitions and results apply for the case $M = 1$.

Definition 4.6. A set $\alpha \in Q_{p,N}$ is *contiguous* if it is related by a cyclic shift to the set $\{0, 1 \ldots, p-1\}$.

Definition 4.7. A set $\alpha \in Q_{p,N}$ is *maximally spread* if it is related by a cyclic shift and/or mirror image to the set $\left\{ \left\lfloor 0\frac{N}{p} \right\rfloor, \left\lfloor 1\frac{N}{p} \right\rfloor, \ldots, \left\lfloor (p-1)\frac{N}{p} \right\rfloor \right\}$.

Conjecture 4.9. *Let $\alpha, \beta \in Q_{p,N}$. If α is maximally spread and β is contiguous, then α is the unique solution to:*

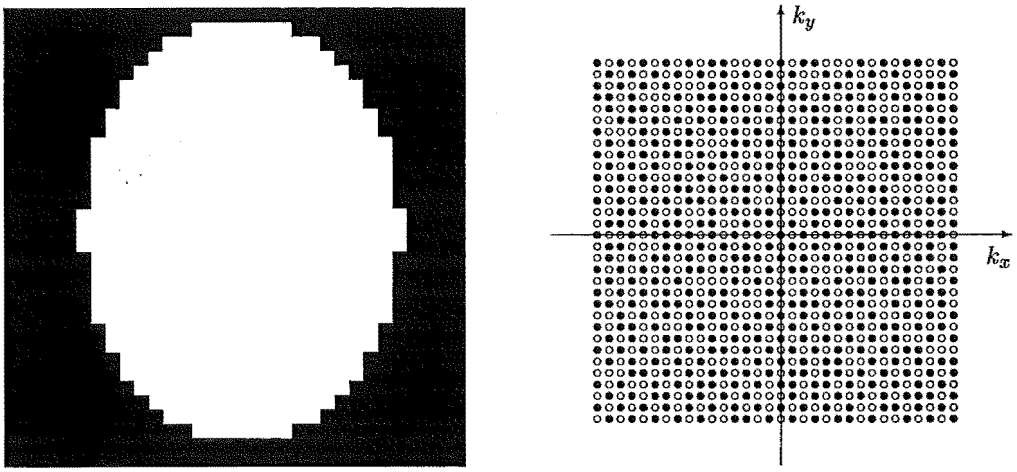
$$\alpha = \underset{\alpha' \in Q_{p,N}}{\operatorname{argmin}} \ \operatorname{tmetric}(W_N[\alpha'|\beta]). \tag{4.53}$$

Note that when p divides N and if the 'unique' qualifier is omitted, the conjecture is true:

Theorem 4.10. *Let $\alpha, \beta \in Q_{p,N}$, where $p \mid N$. If α is maximally spread and β is contiguous, then α satisfies:*

$$\alpha = \underset{\alpha' \in Q_{p,N}}{\operatorname{argmin}} \ \operatorname{tmetric}(W_N[\alpha'|\beta]). \tag{4.54}$$

Proof. Let $A = W_N[\alpha|\beta]$. Recall that $\operatorname{tmetric}(A) = \sum_{i=1}^{p} 1/\sigma_i^2$, and by Theorem 4.6 $\sum_{i=1}^{p} \sigma_i^2 = p^2/N$. The trace metric is therefore minimised if $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_p^2 = p/N$; that is, all eigenvalues of $A^H A$ must be equal to p/N. Since the diagonal elements of $A^H A$ are equal to p/N by Eq. (4.8), this condition is clearly met if $A^H A$ is diagonal. If α is maximally



(a) (b)

Figure 4.3 (a) a 32×32 contiguous 2-D support (white=support), (b) a sampling pattern that yields a well-conditioned reconstruction problem for the support in (a).

spread, then the off-diagonal elements of $A^H A$ are zero, and therefore the trace metric is minimised. □

Calculations over a large number of problem sizes suggest that the conjecture is also true for other values of p with the 'unique' qualifier included.

The natural extension is to consider cases where $M > 1$, but it is not obvious how to extend the notions of contiguous and maximally spread sets. In 2-D, for example, a contiguous set could comprise a single convex region in the 4-connectedness sense (and analogously in higher dimensions), while a maximally spread set could be one in which the spread in the areas of its Voronoi regions were minimised. A fundamental difference between the 1-D case and higher dimensions is that in 1-D there exists only *one* type of contiguous pattern—namely that related to the set $\{0, 1, \ldots, p-1\}$—whereas an M-D definition of contiguous similar to that suggested above would produce a very large number. It is probably unlikely, therefore, that a single maximally spread pattern would be suitable for such a variety of contiguous supports.

Anecdotal evidence in 2-D suggests that for supports comprising a single region, more spread-out patterns generally result in better conditioned problems. For example, Fig. 4.3(a) shows the support of a 32×32 Shepp-Logan head phantom (white=support), and Fig. 4.3(b) gives a sampling pattern that results in very good conditioning, obtained via Gao and Reeves' sample-selection algorithm [GR00] (described in Section 6.2). The measurement locations are clearly dispersed in a manner that seems analogous to the 1-D maximally spread pattern.

4.4.7 Universal patterns

The previous discussion has concerned a particular sampling set being used in conjunction with a particular support, but in some situations it may be desirable to use a *single* sampling pattern for a *variety* of different supports. (In particular, this property becomes essential for the selection of periodic sampling patterns discussed in Section 6.5.) This prompts the question: Is it possible that for a particular problem size, a sampling pattern exists which allows full reconstruction for *any* support? Patterns such as these are called (p, q)-universal [VB00]; as a necessary condition for full reconstruction it must be the case that $p \geq q$. The next question that presents itself is: Will the resulting problems be well-conditioned? Since adding more samples cannot result in worse conditioning (Theorem 4.1), it makes sense to consider the worst case, where $p = q$. These patterns are called (p, p)-universal [VB00], or simply universal:

Definition 4.8. A set $\boldsymbol{\alpha} \in Q_{p,\mathbf{N}}$ is *universal* if it produces a non-singular submatrix of the Fourier matrix for M-D sequences, $W_{\mathbf{N}}[\alpha|\beta]$, for every $\boldsymbol{\beta} \in Q_{p,\mathbf{N}}$. Let $U_{p,\mathbf{N}}$ be the subset of $Q_{p,\mathbf{N}}$ that are universal patterns.

Theorem 4.11. *If $M = 1$, then $|U_{p,N}| \geq 1$ for any p, N.*

Bresler *et al.* [BF96] proved that a 1-D contiguous pattern, α, is universal by exploiting the resulting Vandermonde structure of $W_N[\alpha|\beta]$ [HJ85].

Theorem 4.12. *If $M = 1$ and N is prime, then $U_{p,N} = Q_{p,N}$ for any p.*

This follows immediately from Theorem 4.8, since *any* $p \times p$ submatrix of $W_N[\alpha|\beta]$ is non-singular.

Definition 4.9. Let $\boldsymbol{\alpha} \in U_{p,\mathbf{N}}$. The *best-case trace metric* of $\boldsymbol{\alpha}$ is given by

$$\min_{\boldsymbol{\beta} \in Q_{p,\mathbf{N}}} \text{tmetric}(W_{\mathbf{N}}[\alpha|\beta]). \tag{4.55}$$

Definition 4.10. Let $\boldsymbol{\alpha} \in U_{p,\mathbf{N}}$. The *worst-case trace metric* of $\boldsymbol{\alpha}$ is given by

$$\max_{\boldsymbol{\beta} \in Q_{p,\mathbf{N}}} \text{tmetric}(W_{\mathbf{N}}[\alpha|\beta]). \tag{4.56}$$

These definitions are useful for gauging the performance of a universal pattern. In particular, the worst-case trace metric is henceforth frequently used to compare universal patterns.

Theorem 4.13. *Let $\boldsymbol{\alpha} \in U_{p,\mathbf{N}}$, and let $\boldsymbol{\alpha}' = (\mathbf{a} \circ \boldsymbol{\alpha} + \mathbf{c}) \bmod \mathbf{N}$, where $\mathbf{a}, \mathbf{c} \in \mathbb{Z}^M$, and $(a_j, N_j) = 1$, $j = 0, 1, \ldots, M-1$. Then, $\boldsymbol{\alpha}'$ is universal, i.e. $\boldsymbol{\alpha}' \in U_{p,\mathbf{N}}$, and has the same best-case and worst-case trace metrics as $\boldsymbol{\alpha}$.*

Proof. The two operations (addition and multiplication) may be treated independently, since they can be applied sequentially. Firstly ignoring multiplication (i.e. $\mathbf{a} = \mathbf{1}$), α' is simply a cyclic shift of α. By Theorem 4.3 a cyclic shift does not affect the singular values of $W_{\mathbf{N}}[\alpha|\beta]$, so clearly α' is universal with the same best-case and worst-case trace metrics as α.

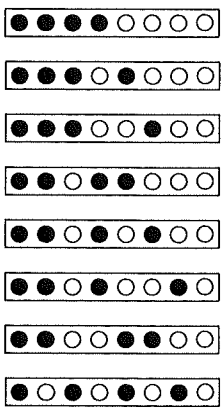
Secondly ignoring addition (i.e. $\mathbf{c} = \mathbf{0}$), note that any $\beta' \in Q_{p,\mathbf{N}}$ can be written in the form $\beta' = (\mathbf{b} \circ \beta) \bmod \mathbf{N}$, where $\mathbf{b} \in \mathbb{Z}^M$, and $a_j b_j \equiv 1 \pmod{N_j}$, $j = 0, 1, \ldots, M-1$. (The existence of β is guaranteed because it can be obtained by $\beta = (\mathbf{a} \circ \beta') \bmod \mathbf{N}$, since a_j and b_j are multiplicative inverses.) $W_{\mathbf{N}}[\alpha|\beta]$ is non-singular for any β since α is universal, and therefore by Theorem 4.4, $W_{\mathbf{N}}[\alpha'|\beta']$ is non-singular for any β'; hence α' is universal.

Let the worst-case trace metric for α be t^{α}_{max} which occurs with β_1, and let the worst-case trace metric for α' be $t^{\alpha'}_{max}$ which occurs with β'_2. Define β'_1 and β_2 such that $\beta'_1 = (\mathbf{b} \circ \beta_1) \bmod \mathbf{N}$ and $\beta'_2 = (\mathbf{b} \circ \beta_2) \bmod \mathbf{N}$. By Theorem 4.4, $\mathrm{tmetric}(W_{\mathbf{N}}[\alpha|\beta_2]) = \mathrm{tmetric}(W_{\mathbf{N}}[\alpha'|\beta'_2]) = t^{\alpha'}_{max}$; the maximum trace metric for α must be at least equal to this by definition, $t^{\alpha}_{max} \geq t^{\alpha'}_{max}$. Similarly, $\mathrm{tmetric}(W_{\mathbf{N}}[\alpha'|\beta'_1]) = \mathrm{tmetric}(W_{\mathbf{N}}[\alpha|\beta_1]) = t^{\alpha}_{max}$ and therefore $t^{\alpha'}_{max} \geq t^{\alpha}_{max}$. Clearly these two constraints are only simultaneously satisfied if $t^{\alpha}_{max} = t^{\alpha'}_{max}$. The proof is identical for the best-case trace metric. □

4.4.8 Applying the theory

The following example illustrates the preceding results, demonstrating their application to a small 1-D sampling system.

Example 4.4. Let $M = 1$, $N = 8$, and $p = q = 4$. The number of possible combinations for a set $\alpha \in Q_{p,N}$ is $\binom{N}{p} = 70$, but after removing those related by a cyclic shift or mirror image (Theorem 4.3), there are only 8 'unique' combinations (see Table 4.1), listed below:



Since α and β can each be any one of these combinations, the entire set of possibilities for

tmetric$(W_N[\alpha|\beta])$ is given by the table below:

$\alpha\downarrow, \beta\rightarrow$	①	②	③	④	⑤	⑥	⑦	⑧
①	241.1	58.6	32.0	32.0	13.4	14.9	16.0	8.0
②	58.6	16.6	18.0	18.0	19.4	13.4	∞	∞
③	32.0	18.0	11.0	∞	18.0	32.0	12.0	∞
④	32.0	18.0	∞	11.0	18.0	32.0	12.0	∞
⑤	13.4	19.4	18.0	18.0	16.6	58.7	∞	∞
⑥	14.9	13.4	32.0	32.0	58.6	241.1	16.0	8.0
⑦	16.0	∞	12.0	12.0	∞	16.0	8.0	∞
⑧	8.0	∞	∞	∞	∞	8.0	∞	∞

Using the theory developed in this chapter, the following results can be determined. They may be confirmed from the above table by the reader.

- Let $\alpha =$ ⑥ and $\beta =$ ②. Application of Theorem 4.4 with $a = 3$ and $b = 3$ gives $\alpha' = \{0,1,2,3\} =$ ① and $\beta' = \{0,3,4,6\}$, which is related to ⑤ by a cyclic shift. Therefore, the entry for (⑥,②) has the same trace metric as that for (①,⑤).

- Let $\alpha =$ ⑦ and $\beta =$ ③. Their complements are $\bar{\alpha} = \{2,3,6,7\}$ and $\bar{\beta} = \{3,4,6,7\}$, which are related to ⑦ and ④, respectively, by cyclic shifts. By Theorem 4.7, the entry for (⑦,③) has the same trace metric as the entry for (⑦,④).

- Let $\beta =$ ④. From the above table, the sampling set that minimises the trace metric is $\alpha =$ ④. By the corollary to Theorem 4.7, the sampling set that minimises the trace metric for the support set $\bar{\beta} = \{2,5,6,7\}$, related to ③ by a cyclic shift, is $\bar{\alpha} = \{2,5,6,7\} =$ ③.

- Let $\beta =$ ①, the contiguous support. By Conjecture 4.9, the unique sampling set that minimises the trace metric is $\alpha =$ ⑧, the maximally spread sampling set.

- From Definition 4.8 and the above table, the universal patterns for this problem size are $\alpha =$ ①,⑥.

- Let $\alpha =$ ①, a universal pattern. Using Theorem 4.13 with $a = 3$ gives $\alpha' = \{0,1,3,6\} =$ ⑥, which is therefore also a universal pattern with the same best-case and worst-case trace metrics.

▲

4.5 Sampling issues

4.5.1 Hermitian symmetry

In the preceding discussion of Sections 4.2 to 4.4, the object was assumed to be complex: $f[\mathrm{n}] \in \mathbb{C}$. If the quantity being imaged is purely real, such as proton density in MRI, then the object's frequency spectrum exhibits Hermitian symmetry: $F[\mathrm{k}] = \overline{F[-\mathrm{k}]}$. Consequently, only one half of the spectrum need be measured to enable full reconstruction of the object. However, if the assumption of Hermitian symmetry is not explicitly incorporated into the reconstruction algorithm, the mere fact that the symmetry exists has no bearing on the reconstruction. In practice, Hermitian symmetry seldom holds exactly, and an incorrect assumption clearly introduces error. Half-Fourier imaging in MRI, where slightly more than half of k-space is usually measured to account for phase errors, was discussed in Section 3.6.2. Measuring half the samples also decreases SNR by a factor of $\sqrt{2}$ compared to full k-space sampling that requires twice as many phase-encoding steps [LL00]. The theory above is entirely compatible with Hermitian sampling, except that some possible sampling sets would be disallowed.

4.5.2 Spread of sampling locations and error measurement

It is interesting to consider the relationship between how spread out a sampling set is and the resulting reconstruction. In Ref. [Fer94] it was shown that for a contiguous 1-D support, the sampling set that gives the best rate of convergence for the iterative algorithms described in the article is where the samples are almost evenly spaced. Conjecture 4.9 suggests that this is also the optimal sampling set for conditioning. But as a consequence of this conjecture, the symmetry of the problem predicts that if the support set is maximally spread, the optimal sampling set is contiguous! It is often assumed that better conditioning is obtained by a more spread-out sampling scheme, however this shows that it is not necessarily the case. And in the 2-D example in Ref. [GR01], it was shown that evenly spread samples can lead to a singular system of equations. The following example facilitates discussion of some of these issues.

Example 4.5. Consider a 1-D cross-section of the Shepp-Logan head phantom (see Fig. 3.9) along the x-axis, shown in Fig. 4.4(a). Let $N = 256$, which gives $q = 177$, and for a slightly over-determined system, let $p = 183$. Consider four sampling sets: 'low-pass', 'high-pass', randomly distributed, and maximally spread. Pseudo-random, complex, white, Gaussian noise was added to the frequency-domain samples to the extent of 30dB SNR, and the reconstructions using Walsh and Nielsen-Delaney's method are shown in Figs. 4.4(b)–(e). Table 4.2 summarises the results, giving root mean square (RMS) error, condition number, trace metric, and the number of significant singular values (NSSV), defined as the number of singular values above the truncation threshold for the reconstruction. ▲

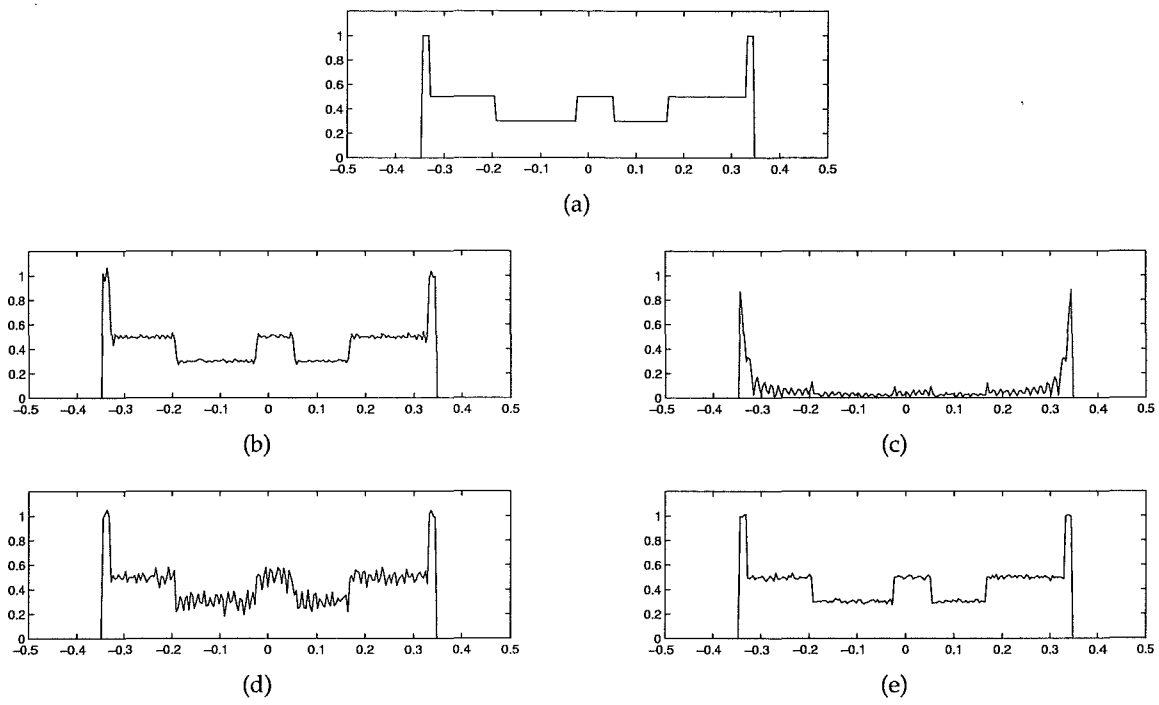


Figure 4.4 Results of Example 4.5: (a) object; reconstructions from (b) low-pass, (c) high-pass, (d) randomly distributed, and (e) maximally spread sampling sets.

Table 4.2 Results of Example 4.5: reconstructions of 1-D signal using four different sampling sets

sampling set	RMS error	cond(·)	tmetric(·)	NSSV
low-pass	0.0209	10^{15}	10^{31}	130/177
high-pass	0.4413	10^{15}	10^{30}	130/177
randomly distributed	0.0661	7,100	10^{7}	170/177
maximally spread	0.0171	2.2	360	177/177

The low-pass and high-pass sampling sets are related by a cyclic shift, meaning the singular values of the two sampling problems are identical (Theorem 4.3). This is reflected in equal condition numbers, trace metrics (apart from machine precision errors), and significant singular values. However, the RMS errors are grossly different. Investigation into the errors due to rank deficiency and noise (see Section 4.4.1) revealed that the errors due to noise, $\|\mathbf{e}_n\|^2$, were approximately equal. This is consistent with equal condition numbers and trace metrics as noted above. On the other hand, the error due to rank deficiency, $\|\mathbf{e}_r\|^2$, was significantly larger for the high-pass sampling set. This can be explained by noting that reconstructions for both the low-pass and high-pass sets had 47 singular vectors omitted, with the aim of reducing the error due to noise at the expense of a larger error due to rank deficiency; how much larger is determined by how much of the signal power is present in the nullspace of the system matrix. Omitting singular vectors effectively enlarges the

nullspace: for the low-pass set there is a small amount of signal power present in this enlarged nullspace, while for the high-pass set there is a sizeable proportion. In short, the properties of the object itself influence the success of the reconstruction.

Another interpretation can be found in the relative magnitude of the object's frequency spectrum: relatively higher magnitude occurs at low frequencies and relatively lower magnitude occurs at high frequencies. As a result, small relative errors on the high-frequency components do not significantly affect the appearance of the reconstruction, but even very small relative errors on the low-frequency components can grossly alter the reconstruction due to their high magnitudes. This explains the large difference in RMS error, since it is *absolute* errors that contribute to RMS rather than *relative* errors.

A question arising from the results above is: Does a larger RMS error correlate with a 'worse' reconstruction? Since the end-user of image reconstruction is usually human, a 'worse' reconstruction would be one that appears worse to the human visual system. A perceptual difference human visual system model was used in Ref. [SDWW99] to compare reconstructed images, with the results comparing well with anecdotal human inspection. The widespread use of RMS error (and the related mean square error) in assessing reconstruction may be due to the fact that RMS error is biased towards components with larger relative magnitude and that these are generally low frequencies; since low frequencies provide the basic structure of an image, it should not be surprising that RMS error correlates reasonably well with human perception.

In addition, although a reconstruction with more accurate low-frequency components may appear generally more like the true object, it may not be very useful if fine detail is required (such as a small brain tumour), which is present in the high-frequency content. The *information content* of different spectral components therefore depends on their relative magnitude, the human visual system, and the task to be performed. Unfortunately, there is no generally accepted method that incorporates these additional features that would enable better assessment of image reconstruction.

The maximally spread sampling set gives the best conditioning as predicted by Conjecture 4.9, and hence the best reconstruction. The randomly distributed sampling set has a condition number and trace metric somewhere between the extremes of the others, but interestingly the reconstruction is slightly worse than for the low-pass set (more singular vectors were retained for the randomly distributed set and this has consequently introduced more error due to noise into the reconstruction). For the contiguous support set, it certainly appears that more spread-out sampling sets result in better conditioned problems. However, the apparently inverse relationship between the spread of the support and the spread of the optimal sampling set is not generally observed for arbitrary support sets.

4.5.3 Conditioning and problem size

Compared to smaller problem sizes, the analysis of larger sizes is characterised by a general increase in complexity due to the increased number of possible combinations of both sampling and support sets. This represents a computational difficulty, but experimental observations suggest that achieving well-conditioned systems for problem sizes with larger N seems more difficult, over and above the extra computational complexity. The following examples investigate this proposition.

Example 4.6. Let $M = 1$ and let $4 \leq N \leq 20$ and $2 \leq q \leq N - 2$ (these combinations of N and q are small enough that unique support sets (see Section 4.4.3) may be exhaustively tested; $q = 1$ and $q = N - 1$ are trivial since for these cases all possible sets are related via a cyclic shift, meaning that there is only one unique set). For each combination of N and q, every unique support set $\beta \in \hat{Q}_{q,N}$ was considered in turn and the minimum trace metric found (such that $p = q$) according to the optimisation

$$\min_{\alpha \in \hat{Q}_{p,N}} \text{tmetric}(W_N[\alpha|\beta]). \tag{4.57}$$

For each combination of N and q, therefore, a total of $\left|\hat{Q}_{q,N}\right|$ minimum trace metrics were obtained. Fig. 4.5 displays the statistics of these populations of minimum trace metrics for all of the problem sizes considered. The minimum, maximum, mean, and median, were computed for each population and are shown as grey levels in (a)–(d), respectively. All trace metrics were first normalised by N to allow fair comparison (see Section 4.4.1). ▲

In Fig. 4.5(a), it can be seen that similar grey levels result for problem sizes with similar q/N ratios. For example, the centre of each row corresponds to $q/N = \frac{1}{2}$, and it is evident that the grey levels appear very similar (and are in fact equal for N even). In other words, for the support sets that have minimum trace metrics equal to the minimum of their respective populations, the trace metric is largely independent of N in absolute terms. However, in Figs. 4.5(b)–(d) a similar pattern is not evident, and instead the grey levels generally increase as N increases. That is, for the support sets that have minimum trace metrics above the minimum of their respective populations, the trace metric depends on the size of N.

These observations suggest the following behaviour: the relative magnitude of the minimum trace metric in Eq. (4.57) depends on β, the ratio q/N, and possibly N. For a 'fortunate' support set, a sampling set may be found to give a trace metric that is equal to the minimum over any combination of sampling and support set for that problem size. Moreover, the relative magnitude of this trace metric is independent of the size of N. For example, a 'fortunate' support set for $N = 10$, $p = 3$ has the same minimum trace metric

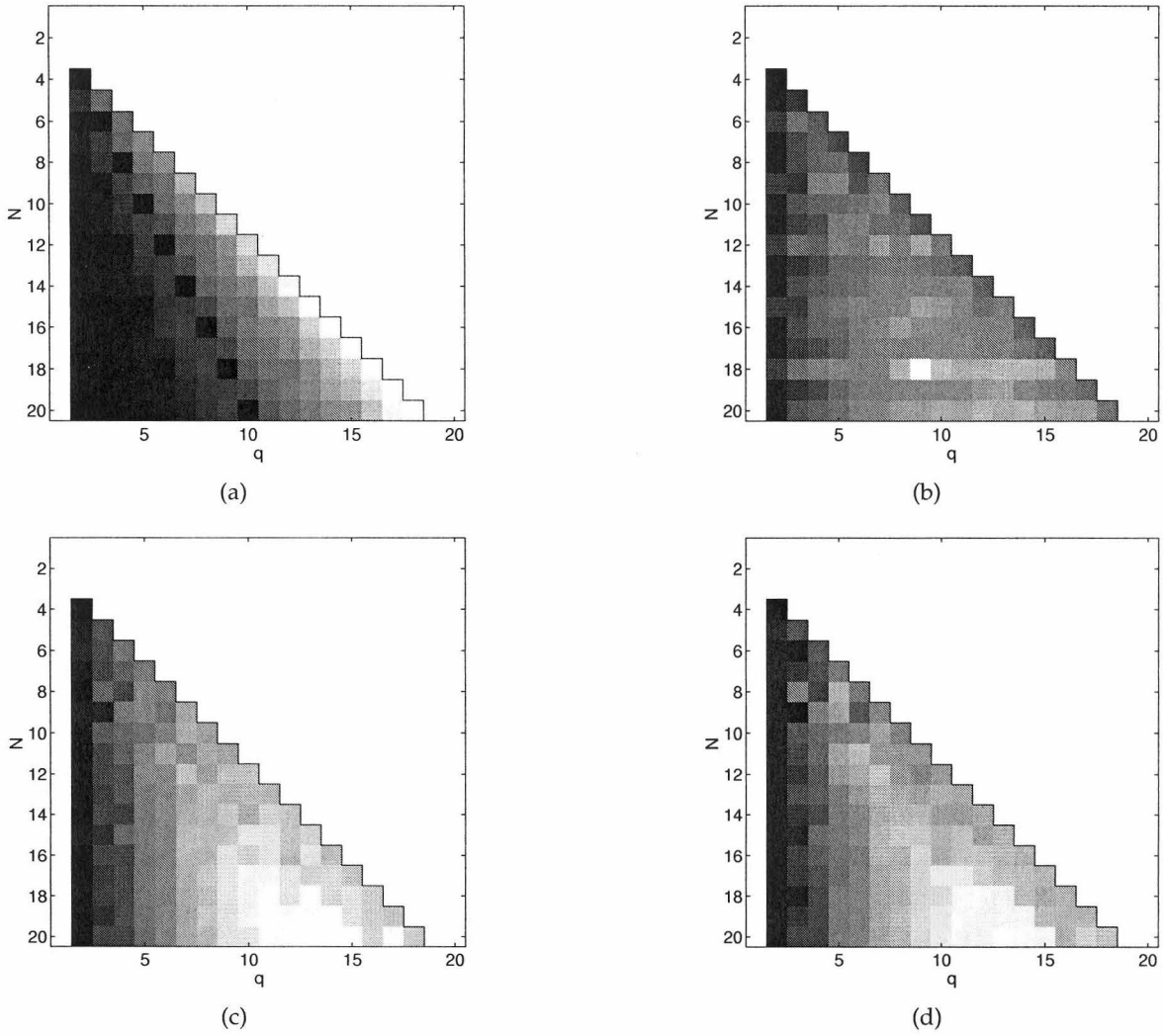


Figure 4.5 Results of Example 4.6: statistics of minimum trace metrics for different problem sizes (combinations of N and q); (a) minimum, (b) maximum, (c) mean, (d) median. Black indicates the lowest value and white the highest value in each plot.

as a 'fortunate' support set for $N = 20$, $p = 6$. However for 'unfortunate' support sets, the minimum trace metric may be comparatively big and will generally be larger if N is larger.

The above example is restricted to reasonably small values of N in order to allow the entire set of combinations to be considered, whereas the following example considers a random subset of combinations for larger values of N.

Example 4.7. Let $M = 1$ and let $4 \leq N \leq 100$ and $2 \leq q \leq N - 2$. For each combination of N and q, 100 random support sets $\beta \in Q_{q,N}$ were considered in turn and the optimisation of Eq. (4.57) was approximated using Gao and Reeves' method [GR00] (see Section 6.2). Therefore, a total of 100 minimum trace metrics were obtained for each combination of N and q (all trace metrics were normalised by N to allow fair comparison). The medians of

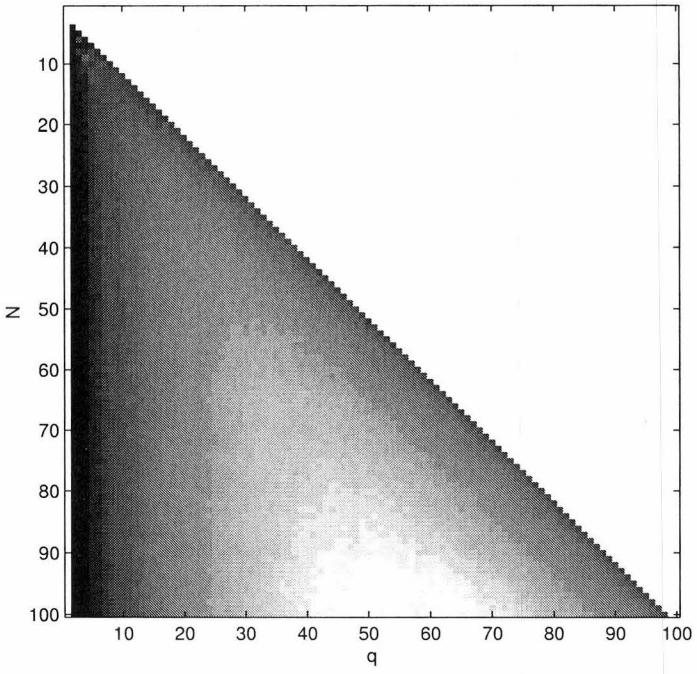


Figure 4.6 Results of Example 4.7: medians of minimum trace metrics for different problem sizes (combinations of N and q). Black indicates the lowest value and white the highest value.

these populations are shown in Fig. 4.6 for all of the problem sizes considered. Because the number of support sets considered as a proportion of the total number of possible support sets (i.e. $100/\left|Q_{q,N}\right|$) differs between problem sizes, it is unfair to compare the minimum and maximum values. The medians are shown as opposed to the means in order to reduce the influence of outliers (although both plots look very similar). ▲

Fig. 4.6 shows that the pattern observed in Fig. 4.5(d) appears to continue for larger N, lending weight to the general behaviour tentatively suggested above. The implication of this behaviour is as follows: Although two problems may appear to be similar in terms of their q/N ratios, it will generally be necessary to set $p > q$ for the problem with larger N to obtain a similarly well-conditioned system.

4.6 Conclusions

A mathematical framework for analysing M-D Cartesian sampling systems has been presented and a body of theory developed that has intuitive, practical, and developmental implications. For example, the result on cyclic shifts and mirror images demonstrates that the modulo nature of a discrete sampling system extends to support and sampling sets, while reducing the number of 'unique' sampling sets that need be considered in practice. In addition, the results on universal patterns and the relationship between 1-D and M-D problems are exploited in Chapter 5 where a sample-selection algorithm is developed. In

light of the theory presented, various sampling issues were discussed: common intuition concerning the spread of the sampling set was challenged and shown to only hold under certain circumstances, while assessing the success of reconstructions was considered from the point of view of the human end user. With the mechanics of Cartesian sampling covered, in the next chapter attention is focussed on the special case of periodicity in the sampling set and its implications for reconstruction algorithms.

Chapter 5

Periodic Cartesian sampling and reconstruction

5.1 Introduction

In the sampling schemes studied in Chapter 4, samples were constrained to lie on a Cartesian grid but otherwise irregularly distributed. What happens if a Cartesian sampling scheme exhibits regularity? Specifically, consider the situation where the overall sampling set comprises a small sampling pattern repeated throughout sampling space. This is the topic investigated in this chapter, where it is shown that periodicity in the sampling set has interesting consequences, with implications for both reconstruction and sample-selection algorithms.

When the sample locations are arbitrary, sampling with a small repeated pattern is called periodic nonuniform sampling. Here, samples are restricted to lie on a Cartesian grid so this type of sampling is referred to as periodic Cartesian sampling (PCS). Fig. 5.1 demonstrates such a sampling scheme, where a small 4×4 block is repeated throughout k-space. Like Cartesian sampling, periodic Cartesian sampling schemes are subsets of the full Nyquist set.

Sampling below the Nyquist density was motivated in Chapter 4 as offering a decrease in imaging time for some MRI techniques, using prior knowledge of the support to enable reconstruction. Periodic Cartesian sampling is simply a special case of this idea and is therefore equally applicable to MR imaging modalities that measure a single point in 2-D k-space per echo. In this context, samples are intentionally omitted to facilitate shorter data acquisition time. However, practical considerations in other sampling procedures may make this type of sampling unavoidable, such as limitations in the design of A/D converters [Jen88].

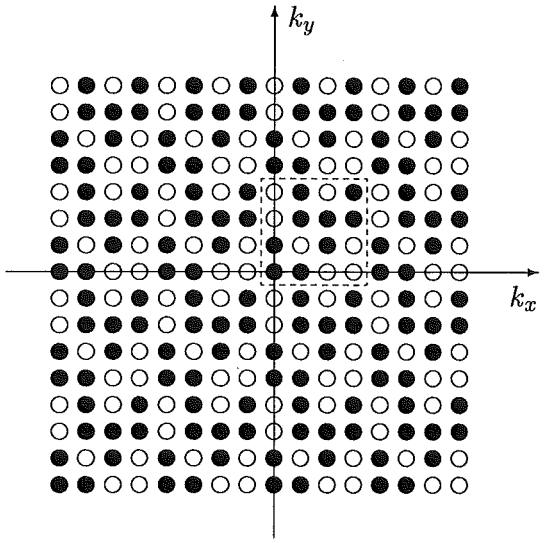


Figure 5.1 Example of a periodic Cartesian sampling scheme: a 4×4 sampling pattern is repeated throughout 2-D k-space; the dotted box indicates the repeated block.

In his paper on irregular sampling, Yen considered periodic nonuniform sampling for the situation of a set of uniform samples migrated to form groups of samples with a recurrent period [Yen56]. Cheung and Marks showed that in some circumstances sampling at the Nyquist density results in linearly dependent sampling due to gaps between the replicated supports in the 'other' domain (i.e. that in which sampling is not performed; see Section 2.2) [CM90]. By exploiting these gaps they demonstrated that some samples can be periodically decimated from a regular grid and reconstructed using an interpolation formula based on the support and deletion geometry. Other authors have also considered how to approach the minimum sampling density via periodic nonuniform sampling, e.g. [VL90, She94a].

Venkataramani and Bresler considered reconstruction from periodic Cartesian sampling, utilising universal patterns (see Section 4.4.7) for sample selection [VB98]. For a 1-D multi-band signal, they investigated how best to achieve the minimum density by optimising the sampling parameters of base sampling rate and sampling pattern [VB00, VB01]. These methods are discussed in more detail later in this chapter and in Chapter 6. The conditions under which full reconstruction is possible for 1-D multi-band signals were studied by Herley and Wong [HW99], and they demonstrated that the minimum rate can be approached or achieved more easily by exploiting these conditions.

MRI was considered explicitly by Nagle and Levin, who used arbitrarily located periodic nonuniform sampling to reconstruct 3-D MR images that have sparse support in multiple regions [NL99, NL01]. Their technique is discussed in more detail later in this chapter. With particular application to MRSI, Gao and Reeves developed a sample-selection algorithm

(described in Chapter 6) for periodic Cartesian sampling [GR01].

This chapter starts by investigating the mathematics of periodic Cartesian sampling. The benefits that emerge lead to the development of a reconstruction technique and a method for predicting the efficacy of a sampling pattern before any data is collected. Following this, additional mathematical analysis is performed to give extra insight into the consequences of periodic Cartesian sampling. Results are presented to demonstrate the developed techniques and finally practical implications are investigated and discussed.

5.1.1 Novel contributions

- **PCSR algorithm** (Section 5.3.2)
  A reconstruction technique for periodic Cartesian sampling appears in Refs. [VB98, VB00, VB01], and the method presented here is essentially the same although it is developed for M-D signals.

- **PCSP algorithm** (Section 5.3.3)
  A method for predicting the success of reconstruction for periodic Cartesian sampling is developed.

- **distribution of singular values among subproblems** (Section 5.3.4)
  An interesting result is presented concerning the distribution of singular values of the overall problem among the subproblems (Theorem 5.2).

- **utility of periodic Cartesian sampling** (Section 5.5)
  The utility of periodic Cartesian sampling is discussed in terms of sampling density and conditioning and it is argued that periodic Cartesian sampling is not disadvantageous compared to general Cartesian sampling.

5.2 Periodic Cartesian sampling

Consider the M-D sampling system developed in Section 4.2 based on discrete sequences defined in Section 2.1. The effects of periodicity in the sampling scheme can be investigated by first considering the general implications of periodicity in a discrete sequence.

Definition 5.1. An M-D discrete sequence $H[\mathbf{k}]$ of size $\mathbf{N}$ is *periodic* if there exists a sequence $G[\mathbf{k}]$ of size $\mathbf{C}$, where $C_j \mid N_j$, $j = 0, 1, \ldots, M-1$, such that

$$H[\mathbf{k} + \mathbf{l} \circ \mathbf{C}] = G[\mathbf{k}], \tag{5.1}$$

where $\mathbf{l} \in \mathbb{Z}^M$.

Define $\mathbf{L} = [L_0, L_1, \ldots, L_{M-1}]^T = \mathbf{N} \circ^{-1} \mathbf{C}$ and let $C = C_0 C_1 \cdots C_{M-1}$ and $L = L_0 L_1 \cdots L_{M-1}$.

Theorem 5.1. *Let $H[\mathbf{k}]$ be an M-D discrete sequence of size $\mathbf{N}$ that is periodic with repeated block $G[\mathbf{k}]$ of size $\mathbf{C}$ and let $\mathbf{L} = \mathbf{N} \circ^{-1} \mathbf{C}$. If the IDFTs of $H[\mathbf{k}]$ and $G[\mathbf{k}]$ are $h[\mathbf{n}]$ and $g[\mathbf{n}]$, respectively, then*

$$h[\mathbf{n}] = \begin{cases} \sqrt{L}\, g[\mathbf{n} \circ^{-1} \mathbf{L}] & \mathbf{n} = \mathbf{m} \circ \mathbf{L},\ \mathbf{m} \in \mathbb{Z}^M \\ 0 & \textit{elsewhere.} \end{cases} \tag{5.2}$$

Proof. Consider the form of $h[\mathbf{n}]$:

$$h[\mathbf{n}] = \frac{1}{\sqrt{N}} \sum_{k_0=0}^{N_0-1} \cdots \sum_{k_{M-1}=0}^{N_{M-1}-1} H[\mathbf{k}] \exp\left[i2\pi \mathbf{k}^T (\mathbf{n} \circ^{-1} \mathbf{N})\right]. \tag{5.3}$$

Letting $\mathbf{k} = \mathbf{c} + \mathbf{l} \circ \mathbf{C}$ gives

$$h[\mathbf{n}] = \frac{1}{\sqrt{N}} \sum_{c_0=0}^{C_0-1} \sum_{l_0=0}^{L_0-1} \cdots \sum_{c_{M-1}=0}^{C_{M-1}-1} \sum_{l_{M-1}=0}^{L_{M-1}-1} H[\mathbf{c} + \mathbf{l} \circ \mathbf{C}] \exp\left[i2\pi (\mathbf{c} + \mathbf{l} \circ \mathbf{C})^T (\mathbf{n} \circ^{-1} \mathbf{N})\right]$$

$$= \frac{1}{\sqrt{N}} \sum_{c_0=0}^{C_0-1} \sum_{l_0=0}^{L_0-1} \cdots \sum_{c_{M-1}=0}^{C_{M-1}-1} \sum_{l_{M-1}=0}^{L_{M-1}-1} G[\mathbf{c}] \exp\left[i2\pi \mathbf{c}^T (\mathbf{n} \circ^{-1} \mathbf{N})\right] \exp\left[i2\pi (\mathbf{l} \circ \mathbf{C})^T (\mathbf{n} \circ^{-1} \mathbf{N})\right]$$

$$= \frac{1}{\sqrt{N}} \sum_{c_0=0}^{C_0-1} \cdots \sum_{c_{M-1}=0}^{C_{M-1}-1} G[\mathbf{c}] \exp\left[i2\pi \mathbf{c}^T (\mathbf{n} \circ^{-1} \mathbf{N})\right] \sum_{l_0=0}^{L_0-1} \cdots \sum_{l_{M-1}=0}^{L_{M-1}-1} \exp\left[i2\pi (\mathbf{l} \circ \mathbf{C})^T (\mathbf{n} \circ^{-1} \mathbf{N})\right]. \tag{5.4}$$

Now let $\mathbf{n} = \mathbf{x} + \mathbf{m} \circ \mathbf{L}$, giving

$$h[\mathbf{n}] = \frac{1}{\sqrt{C}\sqrt{L}} \left\{ \sum_{c_0=0}^{C_0-1} \cdots \sum_{c_{M-1}=0}^{C_{M-1}-1} G[\mathbf{c}] \exp\left[i2\pi \mathbf{c}^T \left((\mathbf{x} + \mathbf{m} \circ \mathbf{L}) \circ^{-1} \mathbf{N}\right)\right] \right\} \left\{ \sum_{l_0=0}^{L_0-1} \cdots \sum_{l_{M-1}=0}^{L_{M-1}-1} \exp\left[i2\pi (\mathbf{l} \circ \mathbf{C})^T \left((\mathbf{x} + \mathbf{m} \circ \mathbf{L}) \circ^{-1} \mathbf{N}\right)\right] \right\}$$

$$= \left\{ \frac{1}{\sqrt{C}} \sum_{c_0=0}^{C_0-1} \cdots \sum_{c_{M-1}=0}^{C_{M-1}-1} G[\mathbf{c}] \exp\left[i2\pi \mathbf{c}^T (\mathbf{x} \circ^{-1} \mathbf{N})\right] \exp\left[i2\pi \mathbf{c}^T (\mathbf{m} \circ^{-1} \mathbf{C})\right] \right\}$$
$$\left\{ \frac{1}{\sqrt{L}} \sum_{l_0=0}^{L_0-1} \cdots \sum_{l_{M-1}=0}^{L_{M-1}-1} \exp\left[i2\pi \mathbf{l}^T (\mathbf{x} \circ^{-1} \mathbf{L})\right] \exp\left[i2\pi \mathbf{l}^T \mathbf{m}\right] \right\}. \quad (5.5)$$

Since $\exp\left[i2\pi \mathbf{l}^T \mathbf{m}\right] = 1$, the second part of this expression is simply the IDFT of a constant, which is a delta function and hence equal to $\sqrt{L}$ when $\mathbf{x} = \mathbf{0}$ and zero elsewhere. Therefore,

$$h[\mathbf{n}]\Big|_{\mathbf{x} \neq \mathbf{0}} = 0, \quad (5.6)$$

and

$$h[\mathbf{n}]\Big|_{\mathbf{x}=\mathbf{0}} = \sqrt{L} \left\{ \frac{1}{\sqrt{C}} \sum_{c_0=0}^{C_0-1} \cdots \sum_{c_{M-1}=0}^{C_{M-1}-1} G[\mathbf{c}] \exp\left[i2\pi \mathbf{c}^T (\mathbf{m} \circ^{-1} \mathbf{C})\right] \right\}$$
$$= \sqrt{L}\, g[\mathbf{m}]. \quad (5.7)$$

With $\mathbf{x} = \mathbf{0}$, $\mathbf{n} = \mathbf{m} \circ \mathbf{L} \Rightarrow \mathbf{m} = \mathbf{n} \circ^{-1} \mathbf{L}$, giving the required result. □

In other words, periodicity in a discrete sequence causes its IDFT to be sparse. This result is illustrated by the following example.

Example 5.1. Consider a 1-D discrete sequence $H[k]$ of size $N = 20$ that is periodic with period $C = 4$, giving $L = 5$:

$$H[k] = [1.9, 0.3, 0.7, 0.3, 1.9, 0.3, 0.7, 0.3, 1.9, 0.3, 0.7, 0.3,$$
$$1.9, 0.3, 0.7, 0.3, 1.9, 0.3, 0.7, 0.3]. \quad (5.8)$$

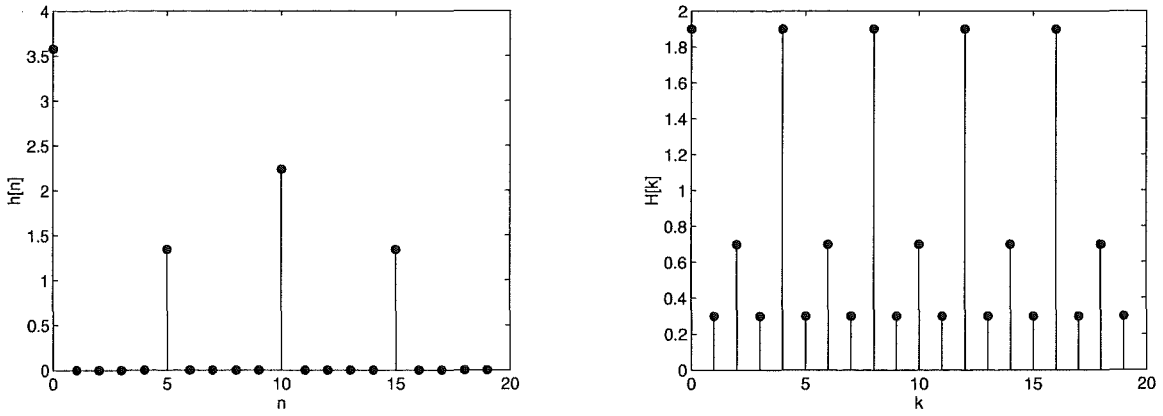
Its IDFT, $h[n]$, is therefore sparse with non-zero values spaced by L, as shown in Fig. 5.2. ▲

Having observed the effects of periodicity in a discrete sequence, the consequences of a periodic sampling pattern are now investigated. The act of Cartesian sampling in general can be viewed as element-by-element multiplication by the sampling mask, $H^{\alpha}[\mathbf{k}]$ (a binary sequence with ones determined by the locations in α; see Definition 4.2). The resulting sequence, $\tilde{F}[\mathbf{k}]$, is therefore

$$\tilde{F}[\mathbf{k}] = F[\mathbf{k}] H^{\alpha}[\mathbf{k}], \quad (5.9)$$

or by the convolution rule,

$$\tilde{f}[\mathbf{n}] = \frac{1}{\sqrt{N}} f[\mathbf{n}] \odot h^{\alpha}[\mathbf{n}], \quad (5.10)$$



(a) (b)

Figure 5.2 Time-domain and frequency-domain sequences for Example 5.1: (a) $h[n]$, (b) $H[k]$.

where $\tilde{f}[\mathbf{n}]$ is the IDFT of $\tilde{F}[\mathbf{k}]$. A single point in the aliased sequence, $\tilde{f}[\mathbf{n}]$, depends on all N points in the original sequence, $f[\mathbf{n}]$, or to put it another way, N points in the original sequence are aliased to a single point. When expressed in this form, the problem is being treated as 'unwrapping' the aliased sequence that results from a reconstruction of just the known samples. Writing the circular convolution as a summation gives

$$\tilde{f}[\mathbf{n}] = \frac{1}{\sqrt{N}} \sum_{l_0=0}^{N_0-1} \cdots \sum_{l_{M-1}=0}^{N_{M-1}-1} f[\mathbf{l}]\, h^{\alpha}[\mathbf{n}-\mathbf{l}]. \tag{5.11}$$

For the case of periodic Cartesian sampling, the sampling mask, $H^{\alpha}[\mathbf{k}]$, is periodic (Definition 5.1). This notion of periodicity can also be expressed in terms of the sampling set:

Definition 5.2. A set $\boldsymbol{\alpha} \in Q_{p,\mathbf{N}}$ is *periodic* with period $\mathbf{C}$ if $\forall\ \mathbf{e} \in \boldsymbol{\alpha}$ it is the case that $[(\mathbf{e} + \mathbf{l} \circ \mathbf{C}) \bmod \mathbf{N}] \in \boldsymbol{\alpha}$, for any $\mathbf{l} \in \mathbb{Z}^M$.

Let the first period of $\boldsymbol{\alpha}$ be $\boldsymbol{\phi} \in Q_{p_0,\mathbf{C}}$ (subsampling set), where $p_0 = p/L$:

$$\boldsymbol{\phi} = \left\{ \mathbf{e} \;\middle|\; \mathbf{e} \in \boldsymbol{\alpha}, e_j < C_j, j = 0, 1, \ldots, M-1 \right\}. \tag{5.12}$$

By Definition 4.2, $H^{\phi}[\mathbf{k}]$ is the mask of $\boldsymbol{\phi}$ and $h^{\phi}[\mathbf{n}]$ is the IDFT of $H^{\phi}[\mathbf{k}]$.

Periodicity in $H^{\alpha}[\mathbf{k}]$ means that $h^{\alpha}[\mathbf{n}]$ is sparse, and consequently the summand in Eq. (5.11)

is only non-zero if $\mathbf{n} - \mathbf{l} = \mathbf{c} \circ \mathbf{L}$, where $\mathbf{c} \in \mathbb{Z}^M$. Hence,

$$\tilde{f}[\mathbf{n}] = \frac{1}{\sqrt{N}} \sum_{c_0=0}^{C_0-1} \cdots \sum_{c_{M-1}=0}^{C_{M-1}-1} f[\mathbf{n} - \mathbf{c} \circ \mathbf{L}]\, h^{\alpha}[\mathbf{c} \circ \mathbf{L}]. \tag{5.13}$$

Now let $\mathbf{n} = \mathbf{a} + \mathbf{m} \circ \mathbf{L}$, giving

$$\tilde{f}[\mathbf{a} + \mathbf{m} \circ \mathbf{L}] = \frac{1}{\sqrt{N}} \sum_{c_0=0}^{C_0-1} \cdots \sum_{c_{M-1}=0}^{C_{M-1}-1} f[\mathbf{a} + (\mathbf{m} - \mathbf{c}) \circ \mathbf{L}]\, h^{\alpha}[\mathbf{c} \circ \mathbf{L}]. \tag{5.14}$$

The notion of a subsequence is now defined:

Definition 5.3. Let $f[\mathbf{n}]$ be an M-D sequence of size $\mathbf{N}$, and let $\mathbf{C}, \mathbf{L} \in \mathbb{Z}^M$ such that $\mathbf{C} \circ \mathbf{L} = \mathbf{N}$. The ith *subsequence* of $f[\mathbf{n}]$ is

$$f_i[\mathbf{m}] = f[\mathbf{a}_i + \mathbf{m} \circ \mathbf{L}], \tag{5.15}$$

where $\mathbf{m} \in \mathbb{Z}^M$ and $\mathbf{a}_i = \mathcal{S}^{-1}(i, \mathbf{L})$, $i = 0, 1, \ldots, L-1$.

In the terminology of multirate digital signal processing [IJ93], a subsequence is simply a decimated (or downsampled) version of the original sequence offset from the origin. The following example demonstrates the relationship between a subsequence and the full sequence.

Example 5.2. A 1-D sequence $f[n]$ of size $N = 20$ has been split into subsequences with the parameters $C = 4$ and $L = 5$. Fig. 5.3 shows $f[n]$, $f_2[m]$, and a block diagram of their relationship. ▲

Finally, Eq. (5.14) may be written in terms of subsequences:

$$\tilde{f}_i[\mathbf{m}] = \frac{\sqrt{L}}{\sqrt{N}} \sum_{c_0=0}^{C_0-1} \cdots \sum_{c_{M-1}=0}^{C_{M-1}-1} f_i[\mathbf{m} - \mathbf{c}]\, h^{\phi}[\mathbf{c}] \tag{5.16}$$

$$= \frac{1}{\sqrt{C}}\, f_i[\mathbf{m}] \odot h^{\phi}[\mathbf{m}], \tag{5.17}$$

where $i = 0, 1, \ldots, L-1$. Thus periodicity has the effect that a point in the aliased sequence is only dependent on the C points of the appropriate subsequence of the original sequence. The problem has essentially been split into L independent and significantly smaller subproblems that must be solved to fully reconstruct the sequence.

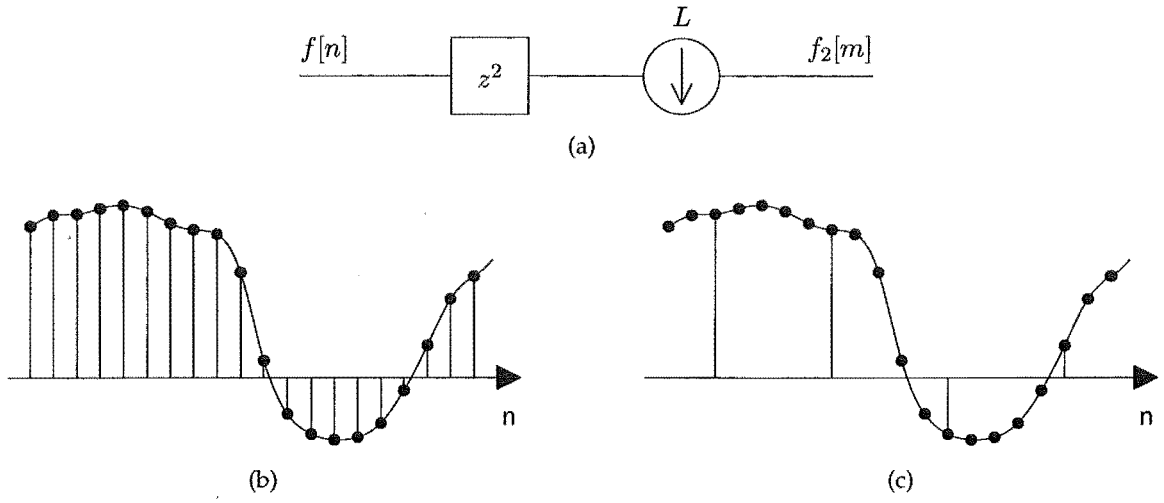


Figure 5.3 The subsequence split described in Example 5.2: (a) block diagram of the process, (b) $f[n]$, (c) $f_2[m]$, for $L = 5$.

5.3 Reconstruction

5.3.1 Basis for reconstruction

Once the problem is stated in the form of Eq. (5.17), a reconstruction technique begins to suggest itself. To view the problem in matrix notation, consider the following definition:

Definition 5.4. The *downsampling matrix*, D_i, is defined as the $C \times N$ submatrix of the $N \times N$ identity matrix, $D_i = I[\gamma^i|\cdot]$, $i = 0, 1, \ldots, L-1$, where

$$\gamma^i = \left\{ \mathbf{e} \;\middle|\; \mathbf{e} = \mathcal{S}^{-1}(i, \mathbf{L}) + \mathcal{S}^{-1}(k, \mathbf{C}) \circ \mathbf{L},\ k = 0, 1, \ldots, C-1 \right\}. \tag{5.18}$$

In 1-D, D_i would simply comprise every Lth row of I starting with row i.

With Definition 5.4, Eq. (5.17) can be written in vector form as

$$\tilde{\mathbf{f}}_i = \frac{1}{\sqrt{C}} \mathbf{f}_i \odot \mathbf{h}^\phi \tag{5.19}$$

$$\Leftrightarrow (D_i \tilde{\mathbf{f}}) = \frac{1}{\sqrt{N}} (D_i \mathbf{f}) \odot (D_0 \mathbf{h}^\alpha), \tag{5.20}$$

where $i = 0, 1, \ldots, L-1$. (The square root factor of N or C depends on whether convolution with $h^\alpha[\mathbf{n}]$ or $h^\phi[\mathbf{n}]$ is considered, respectively, as also seen in Eqs. (5.10) and (5.17); this arises due to the $\sqrt{L}$ factor in Theorem 5.1.) Note that these are M-D circular convolutions. The convolution for the ith subsequence can instead be written as a matrix multiplication (see Section 2.3.3),

$$A^\phi \mathbf{f}_i = \tilde{\mathbf{f}}_i, \tag{5.21}$$

where A^ϕ is the $C \times C$ matrix given by

$$A^\phi = W_{\mathbf{C}}^{-1} \operatorname{diag}(\mathbf{H}^\phi) W_{\mathbf{C}}. \tag{5.22}$$

A^ϕ is circulant in 1-D, block circulant with circulant blocks (BCCB) in 2-D, block circulant with BCCB blocks in 3-D, and so on (see Section 2.3.3 for a discussion of circulant matrices). The elements of A^ϕ are

$$(A^\phi)_{kl} = \frac{1}{\sqrt{C}} h^\phi \left[\mathcal{S}^{-1}(k, \mathbf{C}) - \mathcal{S}^{-1}(l, \mathbf{C})\right]. \tag{5.23}$$

The rank of A^ϕ is equal to the number of measurements in ϕ, i.e. p_0, and is therefore rank deficient by $C - p_0$, meaning that additional information is required in order to fully reconstruct $\mathbf{f}_i$. Knowledge of the support means that some elements of $\mathbf{f}_i$ are known *a priori* to be identically zero; let the support set for the ith subsequence be $\psi^i \in Q_{q_i,\mathbf{C}}$ (subsupport set), where q_i is the number of unknowns. The following example illustrates this concept:

Example 5.3. Let $\mathbf{N} = \binom{30}{30}$ and $\mathbf{C} = \binom{3}{5}$, giving $\mathbf{L} = \binom{10}{6}$. In Fig. 5.4, the grey region represents the region of support within the FOV, and the locations of subsequences 0 and 14 are shown in black in (a) and (b), respectively. It can be verified by observation that the supports for these two subsequences are given by

$$\psi^0 = \left\{ \binom{1}{1}, \binom{2}{1}, \binom{1}{2}, \binom{2}{2}, \binom{1}{3}, \binom{2}{3}, \binom{1}{4}, \binom{2}{4} \right\}, \text{ and} \tag{5.24}$$

$$\psi^{14} = \left\{ \binom{1}{1}, \binom{1}{2}, \binom{2}{2}, \binom{1}{3}, \binom{2}{3}, \binom{1}{4} \right\}. \tag{5.25}$$

▲

The full support, β, can be written in terms of the subsupport sets, ψ^i, as

$$\beta = \bigcup_{i=0}^{L-1} \left[\psi^i + \mathcal{S}^{-1}(i, \mathbf{C})\right], \tag{5.26}$$

and note that

$$\sum_{i=0}^{L-1} q_i = q. \tag{5.27}$$

Incorporating the support information, the problem for the ith subsequence becomes

$$A^\phi[\cdot|\psi^i]\mathbf{f}_i[\psi^i] = \tilde{\mathbf{f}}_i. \tag{5.28}$$

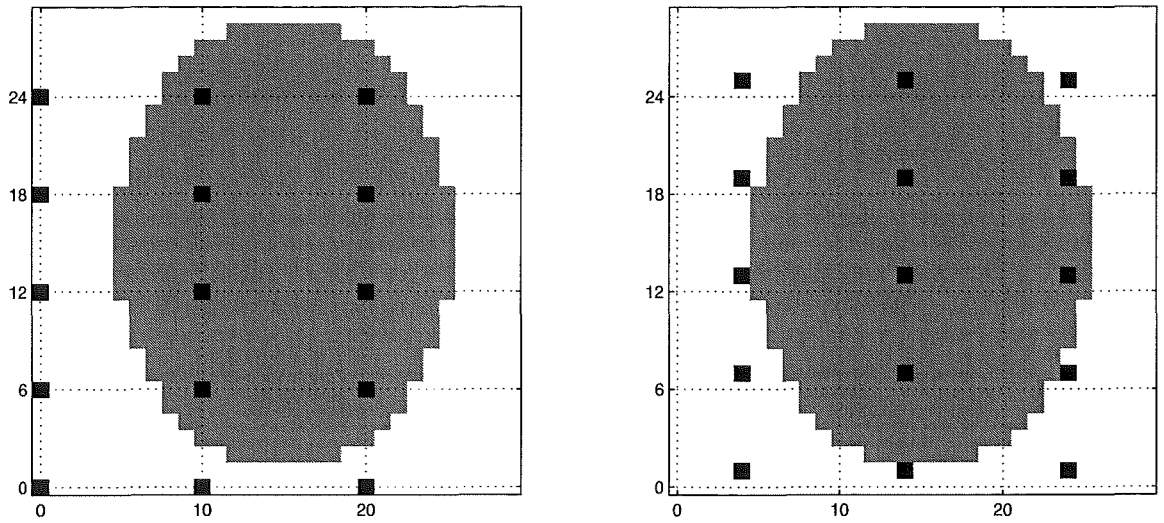


Figure 5.4 Demonstration of subsequences described in Example 5.3: supports for (a) $f_0[\mathbf{n}]$ and (b) $f_{14}[\mathbf{n}]$.

The matrix $A^{\phi}[\cdot|\psi^i]$ has q_i columns, and if $p_0 \geq q_i$ then full column rank (and hence full reconstruction) is possible (though not guaranteed). These subproblems can be solved via the pseudo-inverse (see Section 2.6.1), giving direct, regularised reconstruction and the possibility of obtaining conditioning information for no additional computational effort. The estimated solution, $\hat{\mathbf{f}}_i[\psi^i]$, is given by

$$\hat{\mathbf{f}}_i[\psi^i] = (A^{\phi}[\cdot|\psi^i])^{+}\tilde{\mathbf{f}}_i. \tag{5.29}$$

5.3.2 Reconstruction algorithm

The reconstruction technique outlined above involves splitting the order-N problem into L subproblems, each of order C, which may then be solved independently. To illustrate the practical advantage of this technique, consider the following: the reconstruction of a 256×256 image sampled using a general Cartesian sampling pattern is intractable (given current computer technology) by a matrix-based method that explicitly creates the system matrix due to its enormous size, namely a submatrix containing a large portion of a $256^2 \times 256^2$ matrix. On the other hand, using periodic Cartesian sampling with a repeated 4×4 block results in 64^2 subproblems, each of which only involves a submatrix of a $4^2 \times 4^2$ matrix. Calculation of the pseudo-inverse requires SVD, which has computational complexity of $\mathcal{O}(n^3)$ for an $n \times n$ matrix [GV96], meaning smaller matrices offer a drastic speed improvement (by a factor of 10^{10} in this case). Furthermore, the subproblems are independent and may be solved in parallel if such a computer architecture were available. The reconstruction technique developed here is referred to as PCSR (periodic Cartesian

sampling reconstruction) and is illustrated as a 1-D multirate digital signal processing system in Fig. 5.5. MATLAB® code for the PCSR algorithm appears in Appendix B and it is described algorithmically as follows:

1. Find the periodicity of the sampling set, α, and hence ϕ.
2. Create the matrix A^{ϕ}.
3. For each of the L subsystems:
   (a) Find the support of the subsequence, ψ^i.
   (b) Compute the SVD of $A^{\phi}[\cdot|\psi^i]$.
   (c) Using the singular values, calculate $\mathrm{cond}(A^{\phi}[\cdot|\psi^i])$ and $\mathrm{tmetric}(A^{\phi}[\cdot|\psi^i])$.
   (d) Form the pseudo-inverse (using the SVD) and hence find the minimum-norm least-squares solution to the subproblem (i.e. Eq. (5.29)).

It is worth emphasising the implications of periodic Cartesian sampling for the subsupport sets. While the problem as a whole may be over-determined, fully determined, or under-determined, the subproblems may be any one of these three depending on the particular support constraint and undersampling mask: $p_0 \lessgtr q_i$. It could be the case that for a particular subsequence there are no elements known *a priori* to be zero (i.e. $q_i = C$) in which case $A^{\phi}[\cdot|\psi^i] = A^{\phi}$. Since $A^{\phi} = W_{\mathbf{C}}^{-1}\,\mathrm{diag}(\mathbf{H}^{\phi})W_{\mathbf{C}}$, it is easy to show that A^{ϕ} is Hermitian and idempotent (i.e. $(A^{\phi})^H = A^{\phi}$ and $(A^{\phi})^2 = A^{\phi}$). Consequently, A^{ϕ} is its own pseudo-inverse [Bar90]: $(A^{\phi})^+ = A^{\phi}$. Hence in this special case the minimum-norm least-squares solution is given by

$$\begin{aligned}
\hat{\mathbf{f}}_i[\psi^i] &= (A^{\phi}[\cdot|\psi^i])^+\tilde{\mathbf{f}}_i \\
&= (A^{\phi})^+\tilde{\mathbf{f}}_i && \text{since } A^{\phi}[\cdot|\psi^i] = A^{\phi} \\
&= A^{\phi}\tilde{\mathbf{f}}_i && \text{since } (A^{\phi})^+ = A^{\phi} \\
&= A^{\phi}A^{\phi}\mathbf{f}_i[\psi^i] && \text{by Eq. (5.28)} \\
&= A^{\phi}\mathbf{f}_i[\psi^i] && \text{since } (A^{\phi})^2 = A^{\phi} \\
&= \tilde{\mathbf{f}}_i && \text{by Eq. (5.28).}
\end{aligned} \tag{5.30}$$

The solution via pseudo-inverse is thus simply the original aliased sequence! This makes sense intuitively, since if no additional information is known, no reconstruction should be possible. Interestingly, however, the presence of zeros in the point spread function, $\mathbf{h}^{\phi}$, can produce this result for individual elements of $\hat{\mathbf{f}}_i[\psi^i]$ even if there are known zero elements within the subsequence. Intuitively, zeros in the point spread function mean fewer than C points in the original subsequence are aliased to a single point, and if the points known to be zero fall outside this reduced number of points, the additional information is irrelevant.

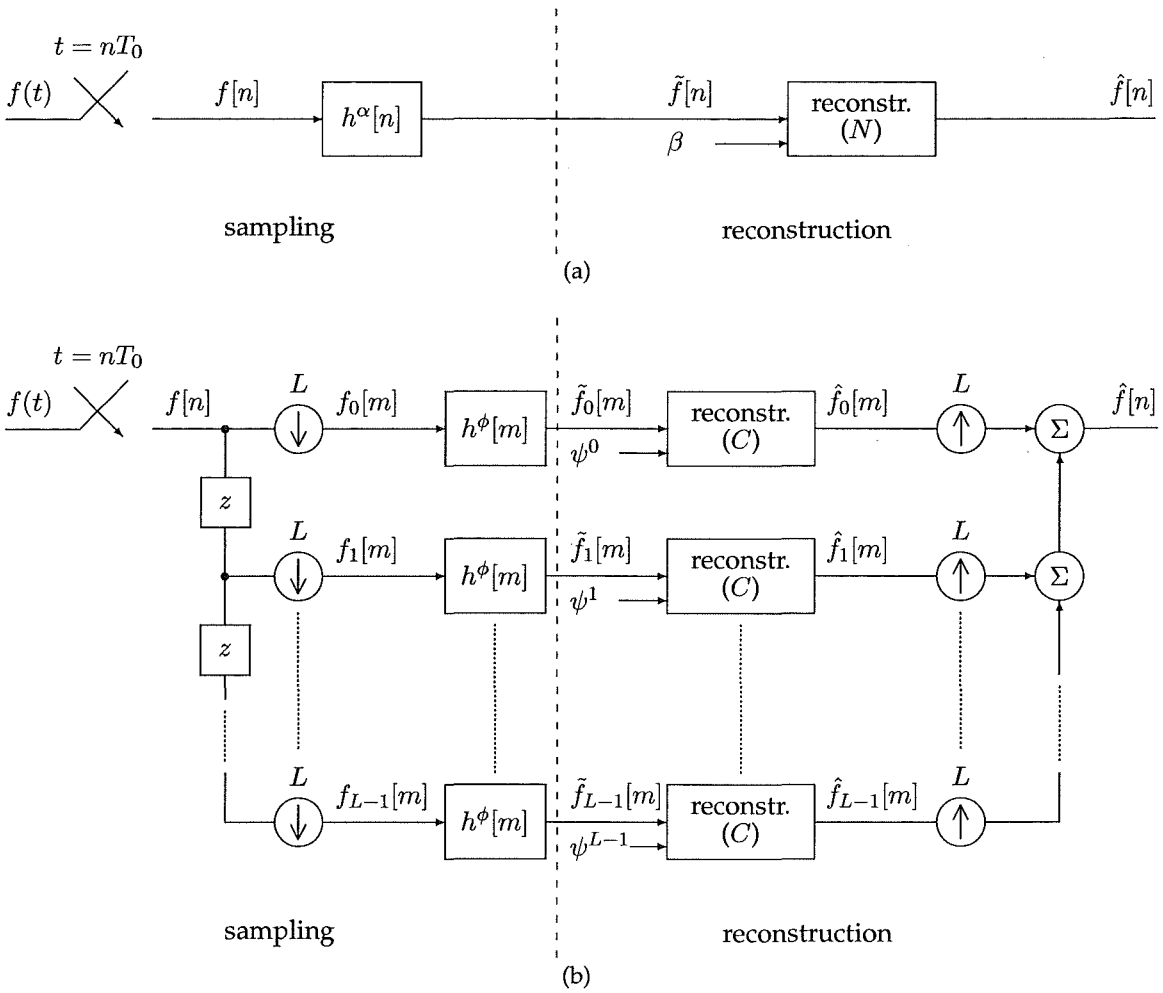


Figure 5.5 Block diagram showing reconstruction with (a) Cartesian sampling (for comparison), and (b) periodic Cartesian sampling (using the PCSR technique described in Section 5.3.2). The reconstruction blocks (marked 'reconstr.') have complexity as shown in parentheses.

At this stage, it is useful to compare the reconstruction technique developed here with others in the literature. The methods of Venkataramani and Bresler [VB98, VB00, VB01] are essentially the same as that developed above: the overall problem is partitioned into subproblems that are solved separately. The PCSR method described here applies to M-D signals that are sampled in the spatial frequency domain and reconstructed in the spatial domain.

Nagle and Levin's multiple region MRI [NL99, NL01] involves splitting the FOV into a relatively coarse grid. If the support is contained in a small number of these grid cells, a periodic nonuniform sampling pattern can be devised to yield reconstruction with very little or no noise amplification. Furthermore, by also sampling a small 'keyhole' in the centre of k-space, they show that reconstruction is possible in the presence of a low-frequency background component not constrained within the support in addition to the high-frequency content inside the support. Fig. 5.6 shows an example where the support comprises three

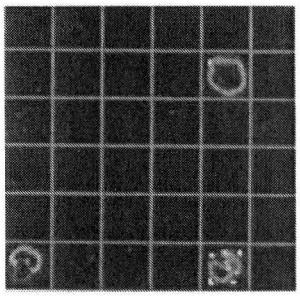


Figure 5.6 Example of a support suitable for multiple region MRI where the signal is only non-zero on three cells of a 6×6 grid covering the FOV [NL99].

cells from a 6×6 coarse grid covering the FOV.

In the above terminology, assuming that complete grid cells are part of the support means that all subproblems have the same support: i.e. $\psi^0 = \psi^1 = \cdots = \psi^{L-1}$. In this way, their method is more restrictive in its assumptions than is PCSR. On the other hand, the sampling pattern is not constrained to lie on a Cartesian grid, although this freedom is a double-edged sword: it allows the creation of sampling patterns that can give very little or no noise amplification, but the selection of such patterns becomes increasingly complex. In Ref. [NL01] the authors propose building a database of suitable patterns for a variety of possible supports that might be encountered; these would need to be computed only once although the computational burden restricts the grids to being relatively coarse. Consequently, their technique is most applicable to situations where the support is covered well by the coarse grid cells and arbitrary *k*-space sampling is possible. This is certainly the case in 3-D MR angiography (MRA), an application suggested by the authors.

5.3.3 Reconstruction prediction

Because PCSR deals with very small matrices, linear algebraic calculations are possible. Furthermore, since the matrix $A^{\phi}[\cdot|\psi^i]$ depends only on the subsampling set, ϕ, and the subsupport set, ψ^i, these calculations can be made *before* any data is collected. For example, to measure the expected performance under noise, the condition number and trace metric could be calculated. Additionally, even if full reconstruction of a subsequence is impossible, it may still be the case that the error is restricted to some regions of the image—it would be useful to know in advance where these regions are.

For any system $A\mathbf{x} = \mathbf{b}$, the general solution consists of the sum of a particular solution and the homogeneous solution. Let A be an $m \times n$ matrix with rank r. There are now r 'basic' variables (those that correspond to columns with leading ones after Gaussian elimination) and $n - r$ 'free' variables (corresponding to columns without leading ones). The free variables are parameters of the homogeneous solution, which forms the nullspace of

A. With $r = n$, there are no free variables and the solution is uniquely determined, but if $r < n$ there are infinitely many solutions [Str88]. However, if a particular basic variable has no dependence on any of the free variables, it is itself uniquely determined.

If the rank of $A^{\phi}[\cdot|\psi^i]$ is less than q_i then full reconstruction is impossible, but some individual elements may still be able to be reconstructed. This can be determined by looking at the nullspace: if the corresponding element in all nullspace vectors is zero, then the basic variable corresponding to that row is uniquely determined. Predicted reconstruction can therefore be established via a two step process:

1. If and only if the rank of $A^{\phi}[\cdot|\psi^i]$ is equal to q_i, then full reconstruction of the ith system is possible;
2. otherwise, if the kth row in all nullspace vectors of $A^{\phi}[\cdot|\psi^i]$ is zero, then reconstruction of $(\mathbf{f}_i[\psi^i])_k$ is possible.

This procedure is referred to as PCSP (periodic Cartesian sampling prediction) and MATLAB® code for the algorithm appears in Appendix B.

Example 5.4. Consider a 1-D discrete sequence $f[n]$ of size $N = 12$ that has support set $\beta = \{0, 1, 2, 3, 7\}$. Its spectrum, $F[k]$, is sampled according to the sampling set $\alpha = \{0, 2, 6, 8\}$, which is periodic with period $C = 6$, giving $L = 2$ and $\phi = \{0, 2\}$. The subsupport sets for the L subsequences are therefore $\psi^0 = \{0, 1\}$ and $\psi^1 = \{0, 1, 3\}$. The masks for all these sets are shown in Fig. 5.7. The matrix $A^{\phi}[\cdot|\psi^0]$ has rank $2 = q_0$ and therefore full reconstruction of the subsequence is possible. However, the matrix $A^{\phi}[\cdot|\psi^1]$ has rank $2 < q_1$ and therefore full reconstruction of this subsequence is not possible. Analysis of the nullspace shows that

$$\mathcal{N}\left(A^{\phi}[\cdot|\psi^1]\right) = \left\{ \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ -\frac{1}{\sqrt{2}} \end{pmatrix} \right\}. \tag{5.31}$$

Since element 1 is zero, reconstruction of $(\mathbf{f}_1[\psi^1])_1$ is possible. If

$$f[n] = [0.95, 0.23, 0.61, 0.49, 0, 0, 0, 0.02, 0, 0, 0, 0], \tag{5.32}$$

then the subsequences are

$$\begin{aligned} f_0[n] &= [0.95, 0.61, 0, 0, 0, 0] \\ f_1[n] &= [0.23, 0.49, 0, 0.02, 0, 0]. \end{aligned} \tag{5.33}$$

Reconstruction via the method of Walsh and Nielsen-Delaney (see Section 4.3) yields

$$\hat{f}[n] = [0.95, 0.12, 0.61, 0.49, 0, 0, 0, 0.12, 0, 0, 0, 0], \tag{5.34}$$

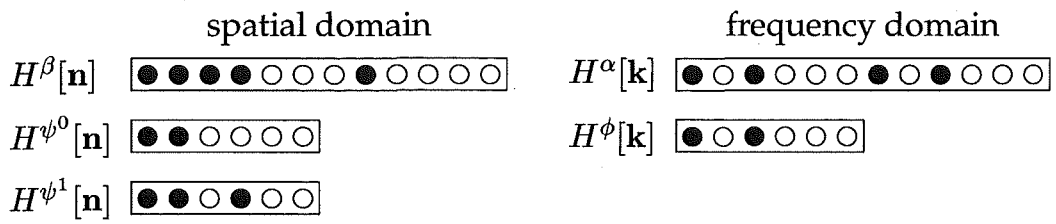


Figure 5.7 Sample and support set masks used in Example 5.4.

giving subsequences of

$$\begin{aligned} \hat{f}_0[n] &= [0.95, 0.61, 0, 0, 0, 0] \\ \hat{f}_1[n] &= [0.12, 0.49, 0, 0.12, 0, 0]. \end{aligned} \tag{5.35}$$

Notice that $f_0[n]$ has been fully reconstructed, while of the three non-zero elements in $f_1[n]$ only the middle element has been reconstructed, as predicted by the above analysis. ▲

Because the prediction depends only on the support and the sampling pattern, it is possible to solve the prediction problem for a particular subsampling pattern and a number of possible subsupports, storing the results in a lookup table (LUT). This would be useful if a particular sampling pattern is to be used a number of times (perhaps one that tends to exhibit good performance in the majority of cases). Then, rather than computing the nullspace of a matrix for each subproblem, a quick check in the LUT would suffice. This would drastically decrease the time required to compute the prediction.

5.3.4 More on subproblems

In Section 5.2, the reconstruction problem was posed in the form of a convolution which lent itself readily to decomposition into subproblems. Here, an alternative approach by linear algebraic methods starting with Eq. (4.12) allows some interesting results to be observed.

Eq. (4.12) states the problem as

$$(E_\alpha W_{\mathbf{N}} E_\beta^T)(E_\beta \mathbf{f}) = (E_\alpha \mathbf{F}), \tag{5.36}$$

where $E_\alpha = I[\alpha|\cdot]$ and $E_\beta = I[\beta|\cdot]$ are extraction matrices (Definition 4.3). Pre-multiplication by $W_{\mathbf{N}}^{-1} E_\alpha^T$ gives

$$(W_{\mathbf{N}}^{-1} E_\alpha^T E_\alpha W_{\mathbf{N}} E_\beta^T)(E_\beta \mathbf{f}) = (W_{\mathbf{N}}^{-1} E_\alpha^T E_\alpha \mathbf{F}). \tag{5.37}$$

Note that the matrix in Eq. (5.37) has the same singular values as that in Eq. (5.36), since

pre-multiplication by E_α^T only introduces zero rows and $W_{\mathbf{N}}^{-1}$ is a unitary matrix. Eq. (5.9) is expressed in vector notation as $\tilde{\mathbf{F}} = \mathbf{F} \circ \mathbf{H}^\alpha$, or equivalently $\tilde{\mathbf{F}} = E_\alpha^T E_\alpha \mathbf{F}$. The aliased sequence is therefore $\tilde{\mathbf{f}} = W_{\mathbf{N}}^{-1} E_\alpha^T E_\alpha \mathbf{F}$. Define $A^\alpha = W_{\mathbf{N}}^{-1} E_\alpha^T E_\alpha W_{\mathbf{N}} = W_{\mathbf{N}}^{-1} \operatorname{diag}(\mathbf{H}^\alpha) W_{\mathbf{N}}$. Analogously to A^ϕ in Section 5.3.1, in 1-D A^α is circulant, in 2-D A^α is BCCB, and so on, defined by the IDFT of $H^\alpha[\mathbf{k}]$, namely $h^\alpha[\mathbf{n}]$ (see Section 2.3.3). Applying these definitions gives

$$(A^\alpha E_\beta^T)(E_\beta \mathbf{f}) = \tilde{\mathbf{f}}. \tag{5.38}$$

Eq. (5.38) is simply an alternative way of expressing the problem in Eq. (5.36), but as seen in Section 5.2, the convolution form of the problem gives insight when the sample set is periodic.

Now assume α is periodic with period $\mathbf{C}$, as developed in Section 5.2. Define the permutation matrix $P = [D_0^T, D_1^T, \ldots, D_{L-1}^T]$, noting that $P^T P = P P^T = I$. Then,

$$P^T A^\alpha P = \begin{bmatrix} D_0 \\ \hdashline D_1 \\ \hdashline \vdots \\ \hdashline D_{L-1} \end{bmatrix} A^\alpha \left[\begin{array}{c:c:c:c} D_0^T & D_1^T & \cdots & D_{L-1}^T \end{array} \right] \tag{5.39}$$

$$= \left[\begin{array}{c:c:c:c} D_0 A^\alpha D_0^T & D_0 A^\alpha D_1^T & \cdots & D_0 A^\alpha D_{L-1}^T \\ \hdashline D_1 A^\alpha D_0^T & D_1 A^\alpha D_1^T & \cdots & D_1 A^\alpha D_{L-1}^T \\ \hdashline \vdots & \vdots & \ddots & \vdots \\ \hdashline D_{L-1} A^\alpha D_0^T & D_{L-1} A^\alpha D_1^T & \cdots & D_{L-1} A^\alpha D_{L-1}^T \end{array} \right]. \tag{5.40}$$

It is shown in Appendix A that

$$D_i A^\alpha D_j^T = \begin{cases} A^\phi & i = j \\ 0 & i \neq j, \end{cases} \tag{5.41}$$

where $A^\phi = W_{\mathbf{C}}^{-1} E_\phi^T E_\phi W_{\mathbf{C}}$, as defined in Eq. (5.22). The matrix $P^T A^\alpha P$ therefore has the form

$$P^T A^\alpha P = \left[\begin{array}{c:c:c:c} A^\phi & 0 & \cdots & 0 \\ \hdashline 0 & A^\phi & \cdots & 0 \\ \hdashline \vdots & \vdots & \ddots & \vdots \\ \hdashline 0 & 0 & \cdots & A^\phi \end{array} \right]. \tag{5.42}$$

That is, $P^T A^\alpha P$ is block diagonal, with each block containing the $C \times C$ circulant matrix A^ϕ.

Rewriting Eq. (5.38) using P gives

$$A^{\alpha} E_{\beta}^{T} E_{\beta} \mathbf{f} = \tilde{\mathbf{f}} \tag{5.43}$$

$$\Leftrightarrow P^{T} A^{\alpha} E_{\beta}^{T} E_{\beta} \mathbf{f} = P^{T} \tilde{\mathbf{f}} \qquad \text{pre-multiply by } P^{T} \tag{5.44}$$

$$\Leftrightarrow P^{T} A^{\alpha} P P^{T} E_{\beta}^{T} E_{\beta} P P^{T} \mathbf{f} = P^{T} \tilde{\mathbf{f}} \qquad \text{since } P P^{T} = I \tag{5.45}$$

$$\Leftrightarrow (P^{T} A^{\alpha} P)(P^{T} E_{\beta}^{T} E_{\beta} P)(P^{T} \mathbf{f}) = (P^{T} \tilde{\mathbf{f}}). \tag{5.46}$$

Since P is unitary, the matrix of Eq. (5.46) still has the same singular values as the matrix in Eq. (5.36). Now letting

$$E_{\psi} = \begin{bmatrix} E_{\psi 0} & 0 & \cdots & 0 \\ 0 & E_{\psi 1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & E_{\psi L-1} \end{bmatrix}, \tag{5.47}$$

it is the case (as shown in Appendix A) that

$$P^{T} E_{\beta}^{T} E_{\beta} P = E_{\psi}^{T} E_{\psi}. \tag{5.48}$$

Eq. (5.46) can now be written in block matrix form:

$$\begin{bmatrix} A^{\phi} & 0 & \cdots & 0 \\ 0 & A^{\phi} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A^{\phi} \end{bmatrix} \begin{bmatrix} E_{\psi 0}^{T} & 0 & \cdots & 0 \\ 0 & E_{\psi 1}^{T} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & E_{\psi L-1}^{T} \end{bmatrix} \cdot \begin{bmatrix} E_{\psi 0} & 0 & \cdots & 0 \\ 0 & E_{\psi 1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & E_{\psi L-1} \end{bmatrix} \begin{bmatrix} D_0 \mathbf{f} \\ D_1 \mathbf{f} \\ \vdots \\ D_{L-1} \mathbf{f} \end{bmatrix} = \begin{bmatrix} D_0 \tilde{\mathbf{f}} \\ D_1 \tilde{\mathbf{f}} \\ \vdots \\ D_{L-1} \tilde{\mathbf{f}} \end{bmatrix}, \tag{5.49}$$

or equivalently,

$$\left[\begin{array}{c|c|c|c} A^{\phi}E_{\psi^0}^T & 0 & \cdots & 0 \\ \hline 0 & A^{\phi}E_{\psi^1}^T & \cdots & 0 \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 0 & \cdots & A^{\phi}E_{\psi^{L-1}}^T \end{array}\right] \left[\begin{array}{c} E_{\psi^0}D_0\mathbf{f} \\ \hline E_{\psi^1}D_1\mathbf{f} \\ \hline \vdots \\ \hline E_{\psi^{L-1}}D_{L-1}\mathbf{f} \end{array}\right] = \left[\begin{array}{c} D_0\tilde{\mathbf{f}} \\ \hline D_1\tilde{\mathbf{f}} \\ \hline \vdots \\ \hline D_{L-1}\tilde{\mathbf{f}} \end{array}\right]. \tag{5.50}$$

Designating the matrix in Eq. (5.50) as B, it is clear that B^HB is block diagonal. Since the determinant of a block diagonal matrix is equal to the product of the determinants of the blocks [HJ85],

$$\det(B^HB - \lambda I) = 0$$

$$\Rightarrow \det\left[E_{\psi^0}(A^{\phi})^HA^{\phi}E_{\psi^0}^T - \lambda I\right] \det\left[E_{\psi^1}(A^{\phi})^HA^{\phi}E_{\psi^1}^T - \lambda I\right] \cdots$$
$$\det\left[E_{\psi^{L-1}}(A^{\phi})^HA^{\phi}E_{\psi^{L-1}}^T - \lambda I\right] = 0. \tag{5.51}$$

The eigenvalues of B^HB are therefore simply the union of the eigenvalues of each block, and hence the singular values of B are the union of the singular values of each block. This result means that investigating properties of the subproblems also gives information about the full problem.

Eq. (5.50) also shows that the problem may be decomposed into L sub-problems,

$$(A^{\phi}E_{\psi^i}^T)(E_{\psi^i}D_i\mathbf{f}) = (D_i\tilde{\mathbf{f}}), \tag{5.52}$$

where $i = 0, 1, \ldots, L-1$. This is equivalently written as

$$A^{\phi}[\cdot|\psi^i]\mathbf{f}_i[\psi^i] = \tilde{\mathbf{f}}_i, \tag{5.53}$$

and comparison with Eq. (5.28) shows that the alternative approach has yielded an identical result.

Finally, pre-multiplication by $E_{\phi}W_{\mathbf{C}}$ gives

$$E_{\phi}W_{\mathbf{C}}W_{\mathbf{C}}^{-1}E_{\phi}^TE_{\phi}W_{\mathbf{C}}E_{\psi^i}^T(E_{\psi^i}D_i\mathbf{f}) = E_{\phi}W_{\mathbf{C}}D_i\tilde{\mathbf{f}} \tag{5.54}$$
$$\Leftrightarrow (E_{\phi}W_{\mathbf{C}}E_{\psi^i}^T)(E_{\psi^i}D_i\mathbf{f}) = (E_{\phi}W_{\mathbf{C}}D_i\tilde{\mathbf{f}}) \tag{5.55}$$
$$\Leftrightarrow (E_{\phi}W_{\mathbf{C}}E_{\psi^i}^T)(E_{\psi^i}\tilde{\mathbf{f}}_i) = (E_{\phi}\tilde{\mathbf{F}}_i), \tag{5.56}$$

where $\tilde{\mathbf{F}}_i = W_{\mathbf{C}}\tilde{\mathbf{f}}_i$. Note that this has the same form as the full problem in Eq. (5.36).

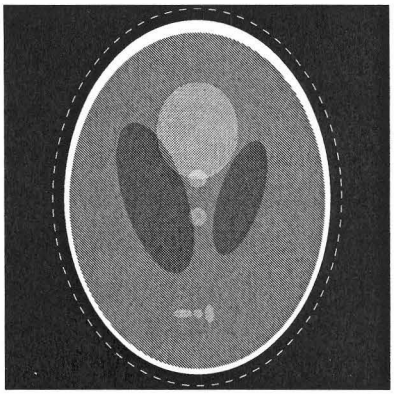


Figure 5.8 The Shepp-Logan head phantom used in the examples; the dashed oval indicates the support that is assumed to be known *a priori*.

The development in this section has hence proved the following theorem:

Theorem 5.2. *Let $\boldsymbol{\alpha} \in Q_{p,\mathbf{N}}$ be periodic with period $\mathbf{C}$ (where $\mathbf{L} = \mathbf{N} \circ^{-1} \mathbf{C}$), and let $\boldsymbol{\beta} \in Q_{q,\mathbf{N}}$. The following are true:*

(a) there exist L independent subproblems, each governed by the matrix $W_{\mathbf{C}}[\phi|\psi^i]$, where $i = 0, 1, \ldots, L-1$; and

(b) the singular values of $W_{\mathbf{N}}[\alpha|\beta]$ are distributed among the subproblem matrices, $W_{\mathbf{C}}[\phi|\psi^i]$.

Corollary. *The trace metric of the overall problem is equal to the sum of the trace metrics of the subproblems.*

Recall that the trace metric is equal to the expected value of the error power in the presence of zero-mean, unit-variance, white, Gaussian noise; intuitively the overall error power should not be altered due to a split into subproblems, as confirmed by the above result.

5.4 Results

Results are presented to illustrate implementation of the reconstruction technique and prediction method developed above. First consider a 240×240 version of the Shepp-Logan head phantom with a relatively loose support (major axis 0.98, minor axis 0.75; see Section 3.8). Fig. 5.8 shows the phantom with the dashed oval indicating the support; this means that 42.3% of the FOV is known *a priori* to be identically zero. Four different sampling patterns were chosen *ad hoc* with repeated blocks of size 2×2, 3×3, 4×5, and 2×6, respectively, shown in the left-hand column of Fig. 5.9. The number of samples per block was chosen so that $p_0/C \approx q/N$, but no attempt was made to ensure full reconstruction was possible; the aim here is to demonstrate aspects of the reconstruction rather than sample selection, with the latter discussed in detail in Chapter 6.

Before samples were 'measured', the support and sampling pattern were used to predict the regions for which reconstruction is and isn't possible ('recoverable' and 'unrecoverable' regions, respectively) and to gather information on the condition numbers and trace metrics of the subproblems. The regions predicted to be unrecoverable are shown in white in the middle column of Fig. 5.9. The sums of the trace metrics for the fully recoverable subsequences were multiplied by the noise power and the results shown in Table 5.1; by Eq. (4.22) and the corollary to Theorem 5.2 these values give the expected value of the error power due to noise in the reconstruction of the fully recoverable subsequences. (The trace metric in any non-fully recoverable subsequences is ∞ by definition so it is only possible to analyse the fully recoverable subsequences, which represent a (usually large) subset of the recoverable regions.)

Finally, samples of the spectrum of the phantom were corrupted with pseudo-random, complex, white, Gaussian noise to the extent of 20 dB SNR and reconstruction was performed via PCSR, with the results shown in the right-hand column of Fig. 5.9. The accuracy of the predictions is evident by observation and this was confirmed numerically by computing the reconstruction without noise and determining those regions identically equal to the 'best estimate', that is, the IDFT of the full Nyquist set (note that this does not correspond to the true object due to the necessary truncation of k-space). Quantitative results are shown in Table 5.1: root mean square (RMS) error and SNR are given for each reconstruction for recoverable (R) and unrecoverable (U) regions along with the overall (O) value, all computed with respect to the true mathematical object. (Note that the grey levels for the phantom cover the range $[0, 1]$, giving RMS error intuitive meaning; see Section 3.8.) The sum of square error (SSE) for the regions corresponding to fully recoverable subsequences (FRS) is also given; this calculation was performed with respect to the 'best estimate' in order to isolate the error due to noise and thereby enable fair comparison to be made with the trace metric sum.

The efficacy of the pre-computed predictions is reinforced by the numerical results: in each case the error predicted by the trace metric is almost identical to the observed value, and the observed SNR in the regions predicted to be recoverable is not substantially lower than the original 20 dB while it is very poor in the regions deemed unrecoverable.

Now consider a 256×256 slice obtained from the 3-D spin warp MRI data set and normalised so that the grey levels cover the range $[0, 1]$ (see Section 3.8), as shown in Fig. 5.10(a) (this is in fact the 'best estimate' of the object since it was obtained by IDFT of the full Nyquist set). As above, a sampling pattern was chosen *ad hoc* with a repeated block of size 4×4 such that $p_0 \approx q/N$, shown in Fig. 5.10(b). PCSP labelled the regions in white in Fig. 5.10(c) as unrecoverable. Reconstructions were then performed on a 1.7GHz AMD^{TM} Athlon PC with 512MB of RAM using both PCSR and the Gerchberg-Papoulis method for

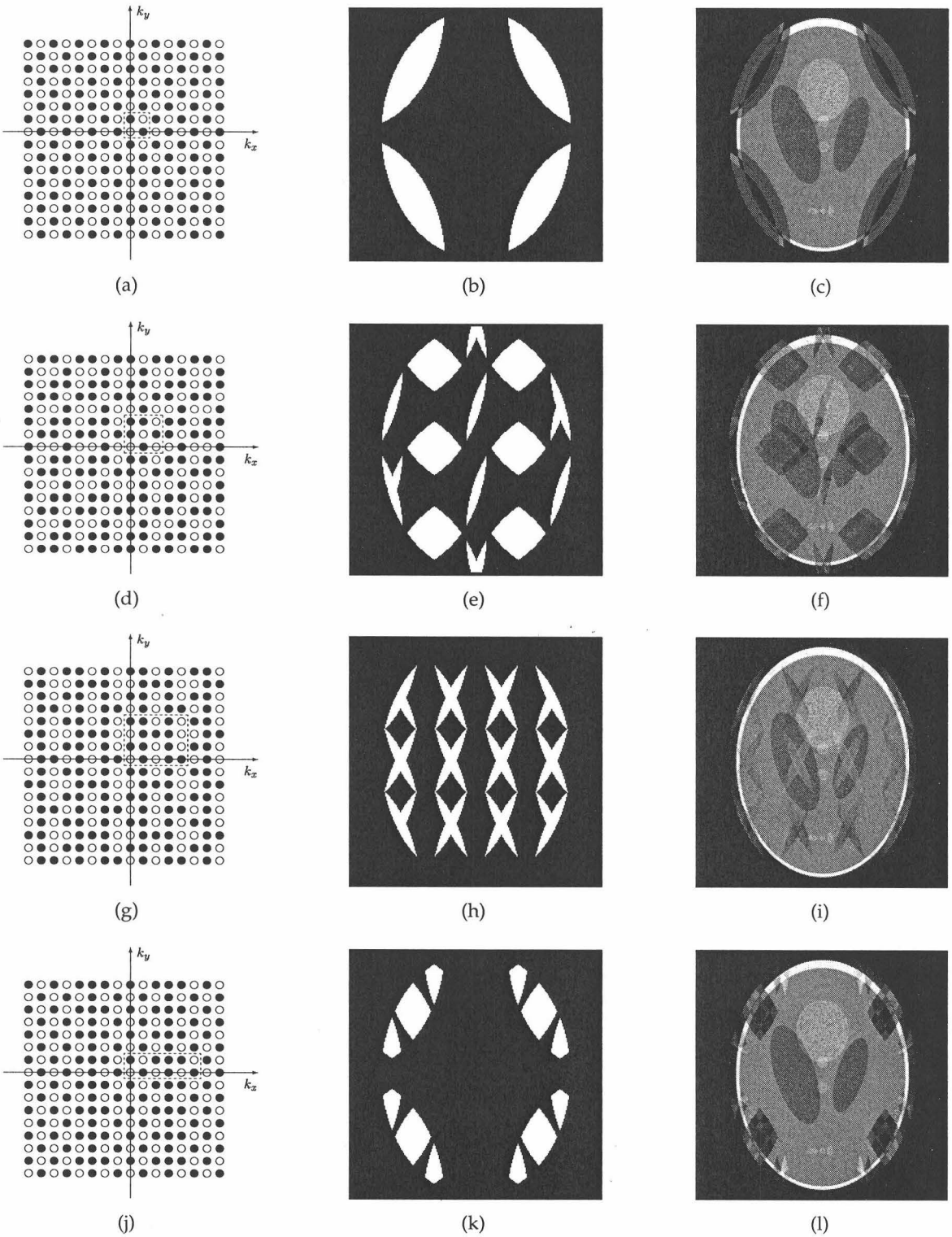


Figure 5.9 Results of the PCSR and PCSP algorithms for the Shepp-Logan head phantom where the samples were corrupted by noise to the extent of 20 dB SNR: (a), (d), (g), (j) 16×16 central portions of k-space indicating the sampling patterns (the dotted lines show the repeated blocks); (b), (e), (h), (k) regions predicted to be unrecoverable (in white); (c), (f), (i), (l) reconstructions from patterns (a), (d), (g) and (j), respectively.

Table 5.1 Numerical results for the reconstructions in Fig. 5.9

sampling pattern	prop. reduc.[a]	sum of tmetric(·)[b]	SSE[c] FRS	RMS error[d] O	RMS error[d] U	RMS error[d] R	SNR (dB)[e] O	SNR (dB)[e] U	SNR (dB)[e] R
1	0.500	62.3	62.3	0.269	0.480	0.063	5.65	0.97	18.09
2	0.444	109.3	109.9	0.235	0.370	0.085	6.81	2.51	15.89
3	0.400	120.7	122.2	0.139	0.228	0.085	11.40	6.32	15.90
4	0.417	70.9	71.0	0.192	0.387	0.070	8.56	3.02	17.16

[a] proportional reduction in sampling density
[b] sum of the trace metrics in the fully recoverable subsequences multiplied by the noise power
[c] sum of square error (SSE) in the fully recoverable sequences (FRS)
[d] root mean square (RMS) error; overall(O) and in the unrecoverable (U) and recoverable (R) regions
[e] signal-to-noise ratio (SNR); overall(O) and in the unrecoverable (U) and recoverable (R) regions

Table 5.2 Numerical results for the reconstructions in Fig. 5.10

reconstruction method	computation time (s)	RMS error[a] O	RMS error[a] U	RMS error[a] R
PCSR	2.2	0.073	0.142	0.059
G-P	20.8	0.073	0.142	0.059

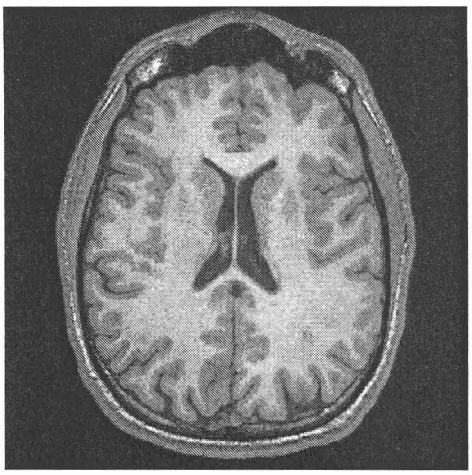
[a] root mean square (RMS) error; overall(O) and in the unrecoverable (U) and recoverable (R) regions

comparison, with the results shown in Figs. 5.10(d) and (e), respectively. The images in Figs. 5.10(a), (d), and (e) are displayed on the range zero (black) to 0.7 (white) in order to enhance the contrast, meaning values above 0.7 are saturated. For the G-P method, the iterations were terminated once successive sum of square errors in the regions outside the support were within 10^{-5}. Table 5.2 gives the computation times and RMS errors.

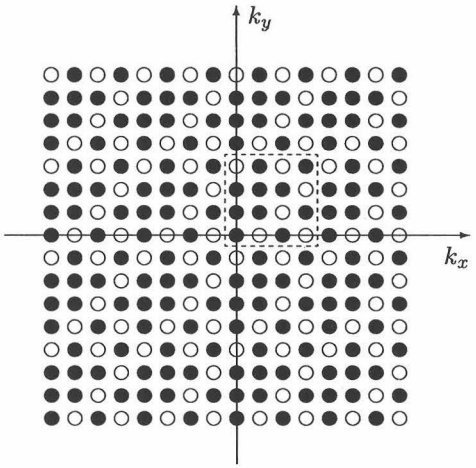
The success of the prediction can be observed in the reconstructed images where the unrecoverable regions appear extremely noisy. It is evident both visually and numerically that the two reconstruction techniques give almost identical results, but the PCSR algorithm took approximately one tenth the computation time while also providing conditioning information.

5.5 Practical implications

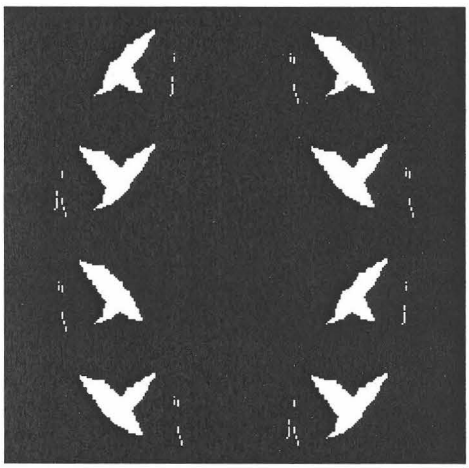
The pivotal difference between Cartesian sampling and periodic Cartesian sampling is the decomposition into subproblems. As a consequence, a *single* subsampling set is used with *multiple* subsupport sets—the success of the reconstruction depends on the performance



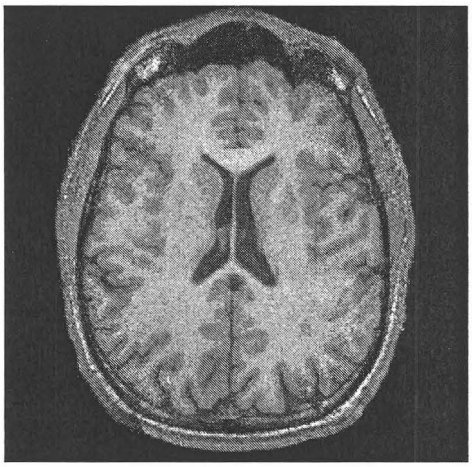
(a)



(b)



(c)



(d)

(e)

Figure 5.10 Results of the PCSR and PCSP algorithms for MRI data: (a) 'best estimate', (b) 16×16 central portion of k-space indicating the sampling pattern, (c) regions predicted to be unrecoverable (in white), (d) reconstruction via PCSR, (e) reconstruction via G-P.

of the subsampling set with each subsupport set. The necessary (though not sufficient) criterion for full reconstruction is also altered: for Cartesian sampling, a fully or over-determined system is required,

$$p \geq q. \tag{5.57}$$

However, for periodic Cartesian sampling a fully or over-determined system is required *for each subproblem*, i.e.

$$p_0 \geq q_i \quad \forall\, i \tag{5.58}$$

$$\Rightarrow p_0 \geq \max_i q_i \tag{5.59}$$

$$\Rightarrow p \geq L \max_i q_i. \tag{5.60}$$

It is necessarily the case that $L \max_i q_i \geq q$, so in general periodic Cartesian sampling requires more samples for full reconstruction than does a Cartesian sampling scheme. However, each subproblem is order C while the overall problem is order N. In Section 4.5.3 it was suggested that for problems with similar q/N ratios, larger problems are generally worse conditioned than smaller problems. This being the case, a slightly oversampled Cartesian sampling scheme may produce similar conditioning performance to a periodic Cartesian scheme selected to exactly determine the subproblem with the largest subsupport, while both give the same reduction in sampling density; this proposition is investigated here. (Note that the term oversampling means sampling above the minimum density, i.e. $p > q$. The object is still undersampled with respect to the Nyquist density.)

Example 5.5. The 2-D support of a 60×60 version of the Shepp-Logan head phantom was considered, meaning 50.1% of the FOV was known *a priori* to be identically zero. The block sizes investigated were 4×4, 5×5, 6×6, 10×10, 15×15, 20×20, 30×30, and 60×60 (note that 60×60 is the special case of no periodicity). For a particular block size, the minimum number of samples per block for full reconstruction, $p_0^{\min}$, was determined using Eq. (5.59), and 51 sampling patterns were chosen with $p_0 = p_0^{\min} + i$, $i = 0, 1, \ldots, 50$, with the constraint that $p < C$. These sampling patterns were selected using Gao and Reeves' technique [GR01] (see Section 6.2), which produces generally good (though suboptimal) results, and the trace metrics for each of the subproblems were subsequently computed.

Cases where $p_0 > p_0^{\min}$ are referred to as block-oversampled, since more samples per block are measured than necessary for full reconstruction. The word 'block' is used to emphasise the difference between oversampling with respect to the minimum density: since it must be the case that $p \geq q$ for PCS, there could be oversampling with respect to the minimum density even when there is no block-oversampling, i.e. $p_0 = p_0^{\min}$. In general, a smaller block size means that the case where $p_0 = p_0^{\min}$ is more oversampled with respect to the minimum density even though there is no block-oversampling.

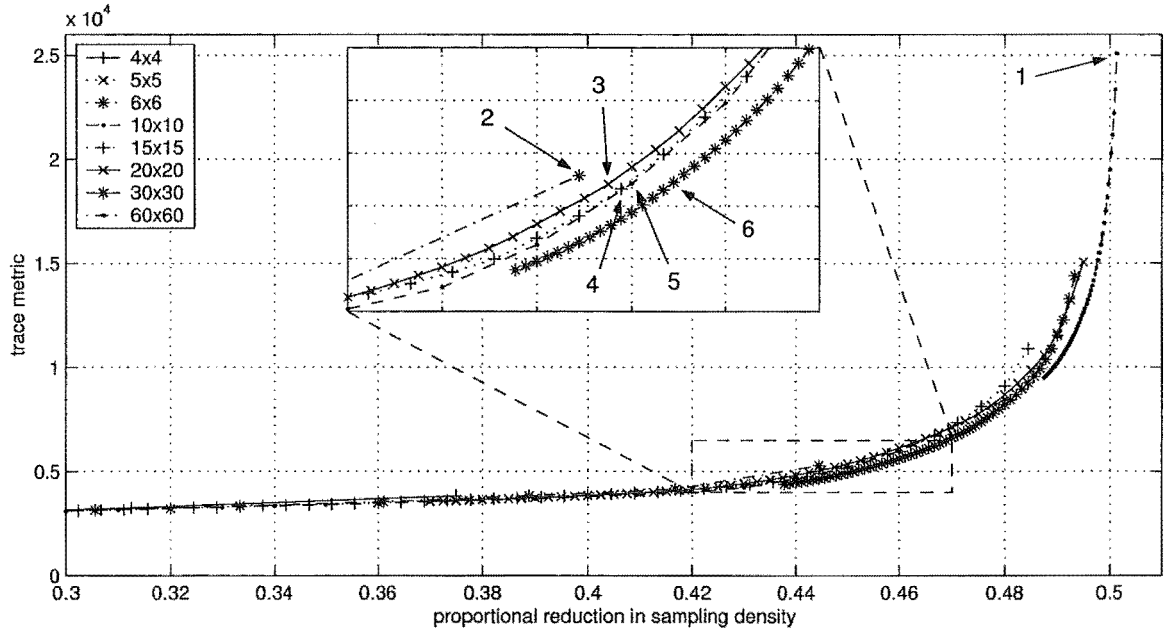


Figure 5.11 Scatter plot for Example 5.5 showing proportional reduction in sampling density and overall trace metric for eight different block sizes and different amounts of block-oversampling.

Fig. 5.11 is a scatter plot of the proportional reduction in sampling density against the overall trace metric, grouped by block size. In each group, the upper-right point represents $p_0 = p_0^{\min}$ and p_0 increases to the lower-left. As expected intuitively, the largest block size ($\mathbf{C} = [60\ 60]^T$, no periodicity) with no block-oversampling gives the best reduction in sampling density ($p = q$), represented by point 1. Oversampling in general provides a tradeoff between sampling density and trace metric: clearly if more samples are measured the sampling density is greater while Theorem 4.1 shows that the error power due to noise cannot increase with more measurements. These characteristics give rise to the downward left shaped curves emanating from the upper-right points representing no block-oversampling. What is also noticeable is the fact that the first extra sample added gives a much greater decrease in trace metric than the second, and so on, a phenomenon also noted in Ref. [GR00]: oversampling reduces the error power but is subject to diminishing returns. This result is also evident on a logarithmic scale though it is slightly less pronounced.

In the lower-left corner of the plot it can be seen that the smaller block sizes have yielded smaller reductions in sampling density. This may be explained by the constraint of Eq. (5.59): very small block sizes result in very small subproblems, and the likelihood of one subproblem having a relatively large q_i is much greater.

The dashed box indicates the portion of the plot that is shown in detail. Point 2 represents $\mathbf{C} = [6\ 6]^T$ with $p_0 = p_0^{\min}$, point 3 represents $\mathbf{C} = [20\ 20]^T$ with $p_0 = p_0^{\min} + 19$, point 4 rep-

resents $\mathbf{C} = [15\ 15]^T$ with $p_0 = p_0^{\min} + 8$, point 5 represents $\mathbf{C} = [10\ 10]^T$ with $p_0 = p_0^{\min} + 3$, and point 6 represents $\mathbf{C} = [30\ 30]^T$ with $p_0 = p_0^{\min} + 35$. All five of these combinations of block size and p_0 give roughly the same trace metric and they span only about 1% difference in proportional reduction in sampling density. In this example, therefore, going to a much smaller block size can give similar conditioning for only a slight increase in sampling density. This is evidence to support the proposition that periodic Cartesian sampling schemes with relatively small block sizes can achieve similar conditioning performance to Cartesian sampling schemes for a comparatively small penalty in terms of sampling density. Together with the fact that periodic Cartesian schemes also allow tractable matrix-based reconstruction, prediction, and conditioning information, PCS provides an attractive alternative to ordinary Cartesian sampling.

Note that only schemes that give full reconstruction were considered in the data presented in Fig. 5.11. If $p_0 < p_0^{\min}$ some subsequences would not be recoverable, but as seen in Section 5.3.3 the error would be constrained to particular regions. Contrast this with the case of no periodicity and $p < q$ in which the error would corrupt the entire image. If a small region of interest (ROI) were desired, an undersampled periodic Cartesian sampling scheme could be designed to allow full reconstruction of that ROI. ▲

To summarise: a very small block size does not generally give good reduction in sampling density, while a very large block size (or no periodicity) sacrifices the computational benefits of PCS. Compared to a heavily block-oversampled pattern with a large block size, a slightly block-oversampled pattern with a moderately small block size (i.e. one with a total number of elements between, say, 20 and 100) can produce similar conditioning performance and give computational advantages at the expense of a slightly smaller reduction in sampling density. In addition, the fact that any error is restricted to specific regions means that undersampling with PCS does not necessarily corrupt the entire image.

Having motivated the use of periodic Cartesian sampling, its implications for sample selection are now discussed. Because one sampling pattern is used with a number of supports, it must be selected with consideration to all supports rather than just one as is the case in Cartesian sampling. At first glance this appears to add to the complexity, but in fact the effect is just the opposite. The sampling pattern to be selected is of order C as opposed to N: fewer samples need to be selected from fewer choices, and notwithstanding the fact that multiple supports must be taken into account, a significant reduction in complexity occurs. As suggested here, a subsampling pattern is selected that is 'tailor made' to the specific subsupport sets encountered. An alternative approach exists where the sampling pattern yields non-singular subproblems for any support set with a certain number of unknowns—universal patterns (see Section 4.4.7). The task of sample selection is discussed in depth in the next chapter.

It is also interesting to consider what happens when there is near-periodic Cartesian sampling. Experiments were conducted in which the locations of measurements were perturbed from a periodic Cartesian sampling pattern, although still located on the Cartesian grid. The results indicate that as the measurements become further removed from their periodic locations, the aliasing that marks the unrecoverable regions gradually fades and therefore the reconstruction improves in those regions. At the same time the near-perfect recovery in the recoverable regions is gradually corrupted, until eventually there is no discernable distinction between the recoverable and unrecoverable regions.

5.6 Conclusions

The consequences of periodic Cartesian sampling have been analysed mathematically and reveal an important result: periodicity in the sampling pattern produces sparsity in the point spread function. This essentially splits the problem into a number of smaller independent subproblems that may be solved separately, as exploited by the PCSR algorithm that has been developed. Taking advantage of periodicity makes possible solving a problem by a linear algebraic method that is otherwise intractable. The drastically smaller matrix sizes also give rise to the PCSP technique for predicting which regions of the object are unrecoverable. The mathematical development concerning subproblems elucidates their relationship to the overall problem.

The Shepp-Logan example demonstrates the efficacy of PCSP and in particular shows how the trace metric can give very accurate prediction of the error due to noise in the reconstruction. Compared to the iterative G-P algorithm in the MR data example, PCSR is roughly ten times faster while also yielding information on conditioning. Use of a computer architecture that allowed the parallelism inherent in the algorithm to be exploited would be faster again. Several criteria were suggested in Section 2.6.2 to evaluate a reconstruction technique: The PCSR algorithm has a demonstrated speed advantage over the G-P method and has been shown to be practical for large images due to the small matrices involved. In addition, reconstruction is perfect in the absence of noise and with suitable regularisation in the pseudo-inverse calculation (such as truncated reconstruction) the sensitivity to measurement noise can be minimised.

The primary argument for using periodic Cartesian sampling over ordinary Cartesian sampling is practical: both sample selection and reconstruction are significantly reduced in complexity. The minimum number of samples required for PCS is at least as large as for CS (Eq. (5.59)), but Example 5.5 shows that once conditioning is taken into account, PCS does not in fact suffer a significant penalty. The phenomenon of recoverable and unrecoverable regions is unique to PCS and offers the flexibility to undersample while restricting the error to well-defined regions that may be predicted in advance. Consequently, a small

ROI could be imaged using an appropriate undersampled PCS scheme. While it is possible in MRI for a particular region to be selectively excited and imaged without the surrounding tissue contributing to the signal, this is not necessarily a simple procedure; the above approach offers an alternative that may also be applicable to other situations.

The sampling patterns chosen *ad hoc* in this chapter give varied results in terms of recoverability—it would be desirable to use a pattern that enables full reconstruction. This task of sample selection is non-trivial and is the subject of the next chapter for both Cartesian and periodic Cartesian sampling schemes.

Chapter 6

Cartesian sample selection

6.1 Introduction

The Cartesian and periodic Cartesian sampling schemes discussed in Chapters 4 and 5 utilise subsets of the full Nyquist set of samples; this chapter considers the task of determining *which* subset should be measured. In other words, the sampling set α is to be found that results in the 'best' reconstruction for a given a support set β. As discussed in Section 2.6.2, the locations of the k-space samples affect (a) whether full reconstruction is possible, and (b) the sensitivity of the reconstruction to noise. The choice of sample locations therefore plays a crucial role in determining the utility of the sampling system; a poor choice may result in such a badly corrupted image that it is essentially useless. Unfortunately, sample selection is a difficult combinatorial problem: there are $\binom{N}{p}$ possible subsets to consider for a given problem size, making exhaustive search impractical for all but the smallest of problems. Furthermore, p may be a parameter of the selection process (providing a tradeoff between sampling density and conditioning) and in PCS the block size must also be chosen.

Having painted the sample-selection problem as dauntingly complex, this chapter begins by surveying some of the current methods of solution. The main drawback of these methods of sample selection is a dependence on linear algebraic techniques to solve the optimisation problem, resulting in computation times that are still impractically long. The aim of this chapter is to develop selection techniques that run considerably faster by using heuristically-based metrics, while still producing well-conditioned systems. Cartesian sampling and periodic Cartesian sampling are dealt with in turn, with results presented after each to demonstrate the efficacy of the techniques developed.

6.1.1 Novel contributions

- **the role of factors** (Sections 6.3.3 and 6.5.1)
  It is argued that singular systems are caused by distances between sample locations that are factors of the size of the discrete sequences. Heuristically-based metrics based on this contention are developed (Eqs. (6.10) and (6.12)).
- **alternative measures of conditioning** (Sections 6.3.2 and 6.5.2)
  Alternative measures of conditioning are proposed that are fast to compute, leading to the development of heuristically-based metrics (Eqs. (6.9) and (6.13)).
- **CSS algorithm** (Section 6.3.4)
  A fast, sequential-selection algorithm is presented for Cartesian sampling.
- **PCSS algorithm** (Section 6.5.3)
  A fast, sequential-selection algorithm is presented for periodic Cartesian sampling.

6.2 Existing methods

Reeves and Heck set up the selection problem as a choice of rows from the matrix modelling the sampling system [RH95]. In the terminology of this thesis, this amounts to selecting the best p rows from an $N \times q$ matrix, $A \in \mathbb{C}^{N \times q}$. They proposed the optimisation cost function $\text{trace}(A^H A)^{-1}$, which is equivalent to $\text{tmetric}(A)$ when A has full column rank (see Section 4.4.1). Since explicit consideration of all possible combinations of rows (a brute-force approach) is computationally intractable, they considered two methods: branch-and-bound, and sequential backward selection (SBS). The former is optimal but computationally expensive, while the latter offers a tradeoff between optimality and computation time. Their SBS algorithm proceeds by sequentially selecting a row to remove until only p rows remain; the row to be removed is the one that increases the cost function the least out of all remaining rows. While they proposed an update formula to simplify computation, the matrix $(A^H A)^{-1}$ must still be stored at each stage of the algorithm, which becomes problematic for large images.

Instead of sequentially eliminating the worst sample location with SBS, Gao and Reeves proposed selecting the best sample using sequential forward selection (SFS) [GR00]. This has the advantage that samples may be measured as soon as they are selected, making real-time selection more feasible. They modified the above cost function to incorporate the case where A is under-determined, yielding $\text{tmetric}(A)$ (recall that this metric is equivalent to the expected value of the reconstruction error power due to noise when the measurements are corrupted with zero-mean, unit-variance, additive, white, Gaussian noise; see Section 4.4.1). With MRSI as the suggested application, matrix A is a submatrix of the

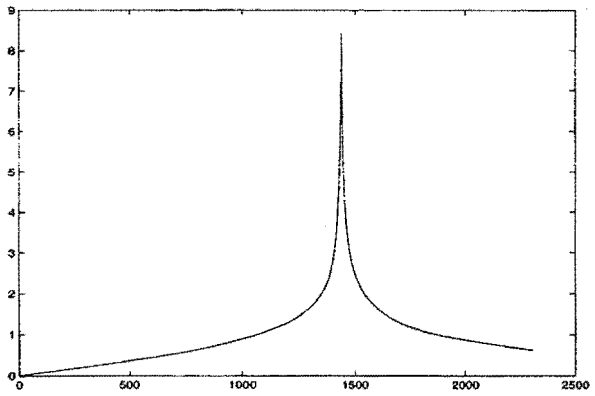


Figure 6.1 Cost function $\mathrm{tmetric}(A)$ plotted against the number of selected samples (p) [GR00].

Fourier matrix and the possible sample locations are from the full Nyquist set, i.e. a Cartesian sampling scheme. To cope with the very large matrices involved and to economise on computation, the authors increased the efficiency of linear algebraic techniques. In particular, matrix inversions are achieved via the conjugate gradient method with the matrix multiplication step computed using FFTs (since A is a submatrix of the Fourier matrix). Selection is still a time-consuming process despite these modifications, taking over an hour for the 64×64 example presented [GR00].

The authors also plotted the value of their selection cost function, $\mathrm{tmetric}(A)$, against the number of samples selected, as shown in Fig. 6.1. Notice that the cost function rapidly decreases immediately after the peak (where the matrix is exactly determined). This means that increasing p when p is only slightly greater than q produces a large marginal decrease in the expected value of the reconstruction error power, but the marginal decrease reduces for larger p. Note that the general behaviour of a marginal increase in $\mathrm{tmetric}(A)$ when $p < q$ and a marginal decrease when $p > q$ is predicted by Theorem 4.1, and the phenomenon of diminishing returns was observed in Example 5.5. Recall too that the perhaps counterintuitive low values of $\mathrm{tmetric}(A)$ for $p < q$ (under-determined problem) are because this measure accounts for the reconstruction error due only to input noise, not that due to under-determinacy.

The techniques outlined above do not restrict the type of sampling pattern used but they suffer from long computational times. To address this problem Gao and Reeves proposed restricting the sampling patterns to be periodic nonuniform [GR01], i.e. a periodic Cartesian sampling scheme. The size of the repeated block is chosen in advance, with examples shown in the paper for 4×4, 8×8, and 16×16. Using a similar selection criterion to that used in Ref. [GR00], their selection technique essentially considers all subproblems at once, choosing the location in the repeated block that has the best effect on the reconstruction. Linear algebraic techniques were developed to facilitate computation, including update

formulas for the metrics, yielding selection in half a minute for the 128×128 example presented.

As discussed in Section 5.5, periodic Cartesian sampling essentially splits the overall problem into independent subproblems, each of which has its own subsupport set. Gao and Reeves' method involves finding a sampling pattern that is 'tailor made' to the particular subsupport sets encountered. Recall that the necessary condition for full reconstruction is given by Eq. (5.59), namely $p_0 \geq \max_i q_i$, where q_i is the number of unknowns in the ith subsupport set. A universal pattern (see Section 4.4.7) that satisfies this condition would therefore ensure full reconstruction of every subproblem—this was the approach taken by Bresler *et al.* [BF96, FB96, VB98, VB00, VB01]. The problem of 'spectrum-blind reconstruction' was discussed in Refs. [BF96, FB96, VB98], with universal patterns playing an integral role (since in this case the actual subsupport sets aren't known *a priori*).

In Refs. [BF96, FB96], Bresler and Feng suggested finding the optimal universal pattern, α^*, by considering the following optimisation problem:

$$\alpha^* = \underset{\alpha \in Q_{p,\mathbf{N}}}{\operatorname{argmin}} \max_{\beta \in Q_{p,\mathbf{N}}} \operatorname{cond} W_{\mathbf{N}}[\alpha|\beta]. \tag{6.1}$$

Since this problem is "very likely NP-complete" [FB96], the authors suggested a simplified search method: a random search gives n candidate patterns, a greedy-style algorithm finds an estimate of the worst-case condition number, then the best $m < n$ patterns are exhaustively searched and the best one chosen. Venkataramani and Bresler developed three metrics for use in the optimisation problem [VB01], which they solved via exhaustive search and forward selection greedy algorithm in the examples presented.

6.3 Cartesian sample selection

This section considers selection for Cartesian sampling schemes (i.e. no periodicity), as developed in Chapter 4. The mathematics presented almost exclusively considers one-dimensional systems (except where indicated), but by Theorem 4.2 an M-D problem is equivalent to a 1-D problem provided that the elements of the size of the discrete sequences are pairwise relatively prime. This allows application of the methods developed to a 2-D example in the next section. The emphasis is on efficient computation and to this end an important relationship concerning conditioning and sample locations is first discussed.

6.3.1 The role of distances

Theorem 4.3 shows that a cyclic shift and/or mirror image of the sampling set and/or the support set does not affect the singular values of the matrix modelling the problem. Since singular values and conditioning are closely related (see Section 4.4.1), it follows that the

conditioning of a problem is not directly dependent on *absolute* sample locations. Thus only the *relative* positions of samples influence a set's properties. These relative positions are characterised by the *distances between sample locations*.

Consider the following 2×2 example:

Example 6.1.

$$W_N[\alpha|\beta] = \frac{1}{\sqrt{N}} \begin{bmatrix} w_N^{-\alpha_0\beta_0} & w_N^{-\alpha_0\beta_1} \\ w_N^{-\alpha_1\beta_0} & w_N^{-\alpha_1\beta_1} \end{bmatrix}, \tag{6.2}$$

for which the trace metric is

$$\text{tmetric}(W_N[\alpha|\beta]) = 2N \left[1 - \cos\left(\frac{2\pi(\alpha_0 - \alpha_1)(\beta_0 - \beta_1)}{N}\right)\right]^{-1}, \tag{6.3}$$

and it is therefore the differences, or distances, between sample locations that is the relevant characteristic. Note that distances are computed modulo N. The system is singular if and only if the trace metric diverges, i.e.

$$(\alpha_0 - \alpha_1)(\beta_0 - \beta_1) \equiv 0 \pmod{N}. \tag{6.4}$$

▲

Cheung and Marks noted that the conditioning of their reconstruction technique depends three factors, including "the distance among the deleted samples" [CM90] (see also Section 4.3). There are good reasons, then, for paying close attention to the distances between sample locations, which may be defined formally for M-D sequences as follows. For a sample set $\boldsymbol{\alpha} \in Q_{p,\mathbf{N}}$ there are $\binom{p}{2}$ pairs of samples, and the vector distance between samples i and j can be written as $|\boldsymbol{\alpha}_i - \boldsymbol{\alpha}_j|$. However, due to the cyclical nature of the discrete sequences this distance is equivalent to many other distances, including $\mathbf{N} - |\boldsymbol{\alpha}_i - \boldsymbol{\alpha}_j|$. For example, in 1-D two adjacent samples are equivalently characterised as being separated by 1 or $(N-1)$. A distance, $\boldsymbol{\delta}$, is therefore constrained such that $0 \leq \delta_j \leq \lfloor N_j/2 \rfloor$, $j = 0, 1, \ldots, M-1$. Note that in 1-D, δ is a scalar on the interval $1 \leq \delta \leq \lfloor N/2 \rfloor$.

Definition 6.1. Let $\boldsymbol{\alpha} \in Q_{p,\mathbf{N}}$. The *distance* between $\boldsymbol{\alpha}_k$ and $\boldsymbol{\alpha}_l$ is $\boldsymbol{\delta}^{\boldsymbol{\alpha}}_{k,l}$, where

$$(\boldsymbol{\delta}^{\boldsymbol{\alpha}}_{k,l})_j = \begin{cases} |(\boldsymbol{\alpha}_k)_j - (\boldsymbol{\alpha}_l)_j| & |(\boldsymbol{\alpha}_k)_j - (\boldsymbol{\alpha}_l)_j| \leq \lfloor N_j/2 \rfloor \\ N_j - |(\boldsymbol{\alpha}_k)_j - (\boldsymbol{\alpha}_l)_j| & |(\boldsymbol{\alpha}_k)_j - (\boldsymbol{\alpha}_l)_j| > \lfloor N_j/2 \rfloor, \end{cases} \tag{6.5}$$

and $j = 0, 1, \ldots, M-1$.

Definition 6.2. The function $\Delta(\boldsymbol{\alpha}, \boldsymbol{\delta})$ is defined as the total number of distances of value $\boldsymbol{\delta}$ in a sample set $\boldsymbol{\alpha} \in Q_{p,\mathbf{N}}$, i.e.

$$\Delta(\boldsymbol{\alpha}, \boldsymbol{\delta}) = \left| \left\{ \boldsymbol{\delta}^{\boldsymbol{\alpha}}_{k,l} \;\middle|\; \boldsymbol{\delta}^{\boldsymbol{\alpha}}_{k,l} = \boldsymbol{\delta},\ k, l = 0, 1, \ldots, p-1,\ k < l \right\} \right|. \tag{6.6}$$

6.3.2 Approximating the trace metric

As noted above and in Section 4.4.1, there is a close relationship between sensitivity to noise (i.e. conditioning) and the singular values of the system matrix. The expected value of the error power due to noise under zero-mean, unit-variance, additive, white, Gaussian noise is given by $\mathrm{trace}(A^H A)^{-1}$ or $\mathrm{trace}(AA^H)^{-1}$ for $p \geq q$ or $p < q$, respectively, which is equal to the sum of the squares of the reciprocals of the singular values (see Section 4.4.1). A perfectly conditioned matrix (according to both $\mathrm{trace}(A^H A)^{-1}$ and $\mathrm{cond}(A)$) occurs when all singular values are equal.

Now let $A = W_N[\alpha|\beta]$, a submatrix of the Fourier matrix, and recall that for any Fourier submatrix the sum of the squares of the singular values, given by $\mathrm{trace}(A^H A)$, is constant and equal to pq/N (Theorem 4.6). Subject to this constraint, it is easy to see that $\mathrm{trace}(A^H A)^{-1}$ is minimised when all singular values are equal. Unfortunately, the matrix inverse presents a problem for efficient computation. Instead consider the sum of the singular values raised to the fourth power, given by $\mathrm{trace}(A^H A)^2$. Again subject to the sum constraint, this value is minimised when all singular values are equal. It is proposed that since $\mathrm{trace}(A^H A)^2$ exhibits the same general behaviour as $\mathrm{trace}(A^H A)^{-1}$, the former may be used as an approximate measure of conditioning.

Calculation of $\mathrm{trace}(A^H A)^2$ can be made very efficient by considering its mathematical form. The elements of $(A^H A)^2$ are

$$(A^H A A^H A)_{kl} = \sum_{m=1}^{q} \left(\frac{1}{N} \sum_{i=1}^{p} w_N^{\alpha_i(\beta_k - \beta_m)} \right) \left(\frac{1}{N} \sum_{j=1}^{p} w_N^{\alpha_j(\beta_m - \beta_l)} \right), \tag{6.7}$$

and therefore the trace is

$$\begin{aligned}
\mathrm{trace}(A^H A)^2 &= \frac{1}{N^2} \sum_{k=1}^{q} \sum_{m=1}^{q} \left(\sum_{i=1}^{p} w_N^{\alpha_i(\beta_k - \beta_m)} \right) \left(\sum_{j=1}^{p} w_N^{\alpha_j(\beta_m - \beta_k)} \right) \\
&= \frac{1}{N^2} \sum_{i=1}^{p} \sum_{j=1}^{p} \sum_{k=1}^{q} \sum_{m=1}^{q} w_N^{(\alpha_i - \alpha_j)(\beta_k - \beta_m)} \\
&= \frac{1}{N^2} \left[pq(p+q-1) + 4 \sum_{i=1}^{p} \sum_{j=i+1}^{p} \sum_{k=1}^{q} \sum_{m=k+1}^{q} \right. \\
&\qquad \left. \cos\left(\frac{2\pi}{N} (\alpha_i - \alpha_j)(\beta_k - \beta_m) \right) \right].
\end{aligned} \tag{6.8}$$

Note that the summand depends on the distances between sample locations in both the sampling and support sets. The summation consists of $\binom{p}{2} \times \binom{q}{2}$ terms, one for each combi-

nation of difference in α and difference in β, and is thus $\mathcal{O}(p^2q^2)$ to compute. Although this appears computationally expensive, very efficient calculation can be achieved in a sequential algorithm, as discussed below in Section 6.3.4. For convenience, the function $\mu^{TS}(\alpha,\beta)$ (trace squared metric) is defined to be the value of $\mathrm{trace}(A^HA)^2$,

$$\mu^{TS}(\alpha,\beta) = \mathrm{trace}((W_N[\alpha|\beta])^H W_N[\alpha|\beta])^2. \tag{6.9}$$

6.3.3 The role of factors

Empirical observations by the author suggested that singular systems occur primarily when distances between sample locations divide N, that is, when $\gcd(\delta, N) > 1$, or in other words δ is either a factor of N or a multiple of a factor of N. This is reinforced by Theorem 4.8, since N prime (no factors) always produces a non-singular system. It is also evident in Example 6.1 that if the distances between the two α samples and two β samples are x and y, respectively, where $xy \equiv 0 \pmod{N}$, a singular system results. However, a general rule describing when singularity occurs based on the distances between sample locations in α and β has proved elusive.

These investigations therefore suggest the following: a problem in which the sampling and support sets contain more distances that are factors (or multiples of factors) of N is more likely to be singular than one with fewer such distances. The following metric (joint factor metric) is proposed for the factors between a sampling and support set, in which a larger (less favourable) metric is given to those combinations in which each set has more samples separated by distances of the type described:

$$\mu^{JF}(\alpha,\beta) = \sum_{x\in\mathcal{F}} \left(\sum_{i=1}^{\lfloor N/2x \rfloor} \Delta(\alpha, ix) \right) \left(\sum_{j=1}^{\lfloor N/2y \rfloor} \Delta(\beta, jy) \right), \tag{6.10}$$

where $y = N/x$ and $\mathcal{F}$ is the set of all factors of N (excluding 1 and N). In Eq. (6.10), $\sum\Delta(\alpha, ix)$ is the number of distances in α with values equal to the factor x or one of its multiples (recall Definition 6.2). Similarly, $\sum\Delta(\beta, jy)$ is the number of distances in β with values equal to the factor y or one of its multiples. Crucially, it is the case that $xy = N$, so distances of ix in α influence the metric more when there are more distances of jy in β and vice versa. In this way, the metric incorporates the strong relationship between distances in both sample sets. Also note that if a particular distance, δ, in α (or β) is a multiple of k factors of N, then $\Delta(\alpha,\delta)$ is included in the summation k times; for example, if δ is a multiple of two different factors of N, then $\delta = ix$ for two different values of x (and two values of i) and it is hence more likely to contribute to a singular system.

6.3.4 Cartesian sample selection algorithm

Using the above development as a basis, an algorithm is described here for efficient selection of a sampling pattern $\alpha \in Q_{p,N}$ given a support set $\beta \in Q_{q,N}$. As a tradeoff between optimality and computation time, a sequential forward selection algorithm is proposed in which decisions are based on the metrics defined above. Specifically, the joint factor metric is used first since this is designed to eliminate any singular systems, and then the trace squared metric is used as an approximation to the trace metric in order to produce a well-conditioned system.

By the corollary to Theorem 4.7, a problem may be approached by investigating its complement, namely selection of a sampling pattern $\bar{\alpha} \in Q_{N-p,N}$ given a support set $\bar{\beta} \in Q_{N-q,N}$. When $p > N/2$ this can simplify (and speed up) the selection process. It essentially corresponds to converting the SFS algorithm into an SBS algorithm and would be less suitable if the sample locations were to be used as soon as they were selected. On the other hand, the speed of this algorithm may mean the delay is acceptable. In addition to speeding up the algorithm, a complementary problems approach means that any unfavourable decisions due to imperfections in the metrics are not able to propagate as far.

The algorithm proceeds as follows and is described precisely by the pseudo-code in Fig. 6.2. First, if $p > N/2$ then β is converted to $\bar{\beta}$ and p to $(N - p)$ for the purposes of selection. Then, since cyclic shifts and mirror images have no effect on a pattern's performance (Theorem 4.3) the first sample point is selected at the origin. The remaining $(p - 1)$ selections are made sequentially: the set of possible sampling locations is first reduced to those with the minimum joint factor metric, then further reduced to those with the minimum trace squared value, and a final random choice is made if needed. As implemented in the examples for this thesis, the random choice is in fact made arbitrarily to be the sample location with the lowest index that hasn't already been selected. The algorithm is referred to as CSS1 (Cartesian sample selection 1-D).

The sequential nature of the algorithm allows for efficient computation of the two metrics, since only a *change* in the metric is relevant at a particular step in the algorithm. Use of a lookup table (LUT) also enables faster computation, and incorporating these techniques, the computational complexity of the CSS1 algorithm is $\mathcal{O}(p^2 N)$. MATLAB® code for the algorithm appears in Appendix B.

For M-D problems, the support must first be converted to 1-D using the relationship in Theorem 4.2 (this involves using the Chinese remainder theorem and also assuming the conversion in Eq. (4.31); this allows a simple application of Eq. (4.29) to the selected 1-D sampling pattern to yield the appropriate M-D result). The 1-D support can then be applied to the CSS1 algorithm and the result converted to M-D. This parent algorithm is

referred to as CSS (Cartesian sample selection) and MATLAB® code for it is also shown in Appendix B.

6.4 Results for Cartesian sample selection

To facilitate display and readability of the results, several conventions are used throughout this section and also in Section 6.6. First, since cyclic shifts and mirror images do not affect conditioning (Theorem 4.3), the subset $\hat{Q}_{p,N}$ is considered that contains only the unique members of $Q_{p,N}$ (see Section 4.4.3). The number of members of $\hat{Q}_{p,N}$ is given in Table 4.1 for small values of N. Similarly, the set $\hat{U}_{p,N}$ contains the unique members of $U_{p,N}$.

The condition $p = q$ is often assumed for testing purposes. Although cases where $p > q$ can also be considered, they have the effect of biasing the results since non-singular and better conditioned systems are 'easier' to obtain in these cases; setting $p = q$ therefore constitutes an acid test.

For relatively small problem sizes, it is possible to search exhaustively through the possible sample sets. The combination $4 \leq N \leq 22$ and $2 \leq p \leq N - 2$ is used, giving a total of 190 problem sizes. These are henceforth described as 'small problem sizes'. Note that $p = 1$ and $p = N - 1$ are trivial since for these cases all possible sets are related via a cyclic shift, meaning that there is only one unique set.

While a system's conditioning may be assessed by the trace metric (for example), this may not be very informative in isolation. Therefore, η is defined as the ratio of the system's trace metric to the trace metric that would result from the best alternative sampling set, that is

$$\eta(\alpha, \beta) = \frac{\text{tmetric}(W_N[\alpha|\beta])}{\min\limits_{\alpha' \in Q_{p,N}} \text{tmetric}(W_N[\alpha'|\beta])}. \tag{6.11}$$

Since the trace metric corresponds to the expected value of the error power (see Section 4.4.1), the ratio $\eta(\alpha, \beta)$, or simply η (usually expressed in decibels), describes the increased level of noise that results from using the given sampling set rather than the best alternative; that is, it describes relative conditioning. Note that a singular system is represented by $\eta = \infty$. If it is impractical to establish the true best alternative (this occurs for larger problem sizes) then an approximation is sometimes used, in which case a tilde reflects this, i.e. $\tilde{\eta}$. Because the best alternative sampling set is not being used, a possible sampling set could produce a system that is in fact better conditioned than the 'best' alternative, in which case $\tilde{\eta} < 0\,\text{dB}$.

Results are first presented that illustrate the efficacy of the two metrics, before the algorithm itself is demonstrated for both small and large 1-D problem sizes and finally for a 2-D example.

Algorithm *Cartesian-Sample-Selection-1-D*(N, β, p)
Input: The problem order, N; the support set, β; and the number of samples to place, p
Output: A suitable sampling pattern, α

/* Use complementary approach if applicable */
flag $\leftarrow 0$
if $p > N/2$
  then $\beta \leftarrow \bar{\beta}$
    $p \leftarrow (N - p)$
    flag $\leftarrow 1$
/* Initialise sampling set and choices set */
$\alpha \leftarrow \{0\}$
$S \leftarrow \{1, 2, \ldots, N-1\}$
repeat
  $T \leftarrow S$
  /* Reduce choices using $\mu^{JF}(\alpha, \beta)$ */
  $T \leftarrow \underset{e \in T}{\operatorname{argmin}} \left[\mu^{JF}(\alpha \cup \{e\}, \beta)\right]$
  /* Reduce choices using $\mu^{TS}(\alpha, \beta)$ */
  $T \leftarrow \underset{e \in T}{\operatorname{argmin}} \left[\mu^{TS}(\alpha \cup \{e\}, \beta)\right]$
  /* Make a final random choice */
  $e \leftarrow$a randomly selected element of T
  /* Update sets */
  $\alpha \leftarrow \alpha \cup \{e\}$
  $S \leftarrow S \setminus \{e\}$
until $|\alpha| = p$
/* Compute complement if necessary */
if flag $= 1$
  then $\alpha \leftarrow \bar{\alpha}$
return α

Figure 6.2 Pseudo-code for the CSS1 algorithm. The argmin operator returns the set of arguments that produce the function's minimum value.

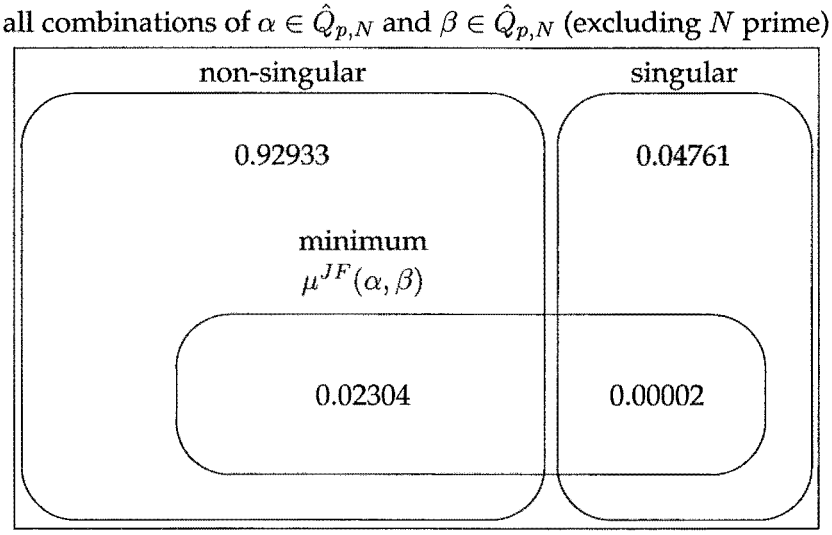


Figure 6.3 Venn diagram showing the success of $\mu^{JF}(\alpha, \beta)$. The numbers indicate fractions of all possible systems (combinations of $\alpha \in \hat{Q}_{p,N}$ and $\beta \in \hat{Q}_{p,N}$) for small problem sizes (excluding N prime).

6.4.1 Performance of the metrics

Joint factor metric

In analysing $\mu^{JF}(\alpha, \beta)$, cases where N is prime were excluded for two reasons: all values of $\mu^{JF}(\alpha, \beta)$ would be zero, and by Theorem 4.8 all such systems are non-singular. For each of the remaining 136 small problem sizes, all possible support sets $\beta \in \hat{Q}_{p,N}$ were considered and the joint factor metric was calculated for every possible sampling set $\alpha \in \hat{Q}_{p,N}$ (note that $p = q$). Systems were classified by whether they provided the minimum joint factor metric and by singularity.

The results are shown as a Venn diagram in Fig. 6.3: the rectangle represents all combinations of $\alpha \in \hat{Q}_{p,N}$ and $\beta \in \hat{Q}_{p,N}$ (i.e. all possible systems) for small problem sizes (excluding N prime) and a number in a particular region indicates the fraction of systems that belong to that region. As may be seen from the diagram, 99.9% of systems that had the minimum joint factor metric were non-singular. Since the sole purpose of this metric is to identify non-singular systems, it can be concluded that it fulfills this role almost perfectly.

Trace squared metric

To evaluate $\mu^{TS}(\alpha, \beta)$, all combinations of $\alpha \in \hat{Q}_{p,N}$ and $\beta \in \hat{Q}_{p,N}$ were considered for all 190 small problem sizes. Systems were classified by whether they produced the minimum trace squared metric, by singularity, and by whether they exhibited $\eta < 3\,\text{dB}$ conditioning.

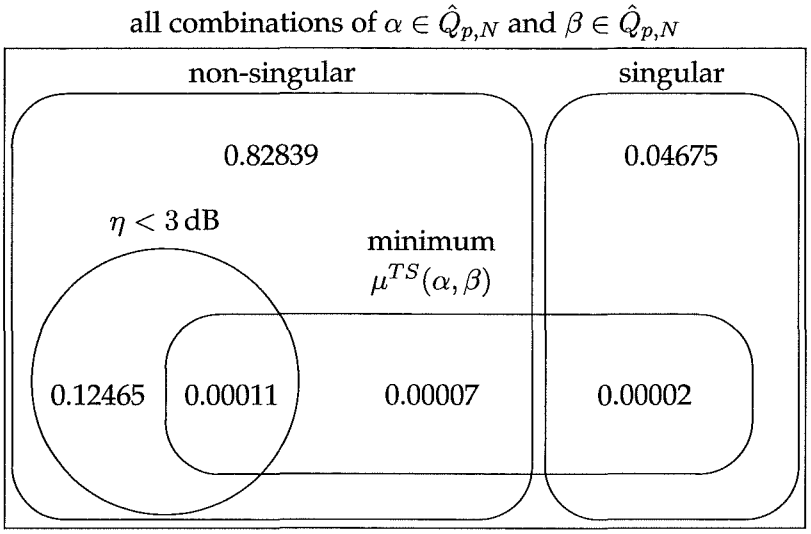


Figure 6.4 Venn diagram showing the success of $\mu^{TS}(\alpha, \beta)$. The numbers indicate fractions of all possible systems (combinations of $\alpha \in \hat{Q}_{p,N}$ and $\beta \in \hat{Q}_{p,N}$) for small problem sizes.

The Venn diagram in Fig. 6.4 shows the results: the rectangle represents all possible systems for small problem sizes and a number indicates a fraction of systems. It is evident from the diagram that 89% of systems that had the minimum trace squared metric were non-singular and 53% had $\eta < 3$ dB conditioning, showing good performance of the metric. Identifying non-singular systems only 89% of the time is certainly not ideal, highlighting the need for an additional singularity-based metric, namely the joint factor metric.

6.4.2 Performance of the algorithm

Small one-dimensional problems

For each of the 190 small problem sizes, the algorithm was applied to every possible support set $\beta \in Q_{p,N}$, giving a total of 206,006 trials. Only 0.06% of the resulting systems were singular, and in the non-singular cases the mean value of η was 14.7 dB. The results are presented in Table 6.1 with systems classified by η: the first row gives the raw numbers of systems in each category of η and the second row expresses the numbers as proportions of the total number of systems tested. The prevalence of systems with low values of η (64.2% had $\eta < 6$ dB) shows that the algorithm produced sampling sets that resulted in well-conditioned systems the vast majority of the time.

Large one-dimensional problems

To evaluate the algorithm's performance for larger problem sizes, 100 sizes were randomly selected such that $23 \leq N \leq 500$ and $2 \leq q \leq N - 2$, and for each size 100 support sets were randomly chosen. The algorithm was run with each support set three times,

Table 6.1 Results of the CSS1 algorithm for small 1-D problems

	range of η (dB)				
	0–3	3–6	6–10	10–∞	∞
number of systems[a]	79,589	52,599	39,166	34,536	116
proportion of systems[b]	0.386	0.255	0.190	0.168	0.001

[a] raw number of systems with relative conditioning η in each category
[b] number of systems with relative conditioning η in each category as a proportion of the total number tested

for $p = q$, $p = q + 10$ (if applicable), and $p = q + 50$ (if applicable), in order to judge the performance with different levels of oversampling. Since finding the best alternative sampling set by exhaustive search is impractical for these problem sizes, Gao and Reeves' SFS algorithm [GR00] was used as the benchmark for $\tilde{\eta}$.

The problem sizes selected resulted in 10,000 systems for $p = q$, 9,600 systems for $p = q+10$, and 7,000 systems for $p = q + 50$. A singular system was produced in only 171 cases for $p = q$ (1.71%), only six cases for $p = q + 10$ (0.06%), and never for $p = q + 50$. The mean values of $\tilde{\eta}$ for the three levels of oversampling (excluding the singular cases) were 43.0 dB, 26.5 dB and 0.0 dB, respectively. The results are shown in Table 6.2, classified by $\tilde{\eta}$: for each level of oversampling, the first row gives the raw numbers of systems in each category of $\tilde{\eta}$ and the second row expresses the numbers as proportions of the total number of systems tested (for that level of oversampling). They seem to suggest that the algorithm does not perform especially well for exactly determined systems, but works much better when the system is over-determined. This may not be a significant disadvantage: since sampling slightly above the minimum density can realise large reductions in the trace metric obtained, it is probably desirable to operate in the slightly oversampled region anyway.

It is interesting to consider why the algorithm does not perform better for exactly or very slightly over-determined systems. In Section 6.4.1, the joint factor metric was shown to identify non-singular systems almost all of the time. However, the proportion of non-singular systems identified was small (i.e. the joint factor metric is very conservative), meaning that the subset of possible sampling sets that are input to the trace squared metric is small. Together with the fact that the trace squared metric is only an approximation to conditioning, this explains the sometimes poor performance of the algorithm. If the joint factor metric could be modified to admit a larger subset of non-singular systems while still eliminating singular systems, the trace squared metric would be able to function more effectively.

Table 6.2 Results of the CSS1 algorithm for large 1-D problems

oversampling[a]		range of $\tilde{\eta}$ (dB)				
		< 3	3–6	6–10	10–∞	∞
0	number of systems[b]	1,201	826	2,183	5,619	171
	proportion of systems[c]	0.12	0.08	0.22	0.56	0.02
10	number of systems	8,038	889	268	399	6
	proportion of systems	0.84	0.09	0.03	0.04	0.00
50	number of systems	6,946	36	11	7	0
	proportion of systems	0.99	0.01	0.00	0.00	0.00

[a] number of samples added above the minimum density
[b] raw number of systems with relative conditioning $\tilde{\eta}$ in each category
[c] number of systems with relative conditioning $\tilde{\eta}$ in each category as a proportion of the total number tested

Two-dimensional example

Consider a 60×61 version of the Shepp-Logan head phantom with a relatively loose support (as used in Section 5.4 and shown in Fig. 5.8 on page 97). The size was chosen such that its elements are relatively prime in order for 1-D and 2-D equivalence to hold. The size was also chosen to make the problem tractable, and it should be noted that the resulting relatively small extent of k-space considered leads to relatively low resolution (even if the full Nyquist set were measured).

The support was applied to the CSS algorithm to select a 2-D sampling pattern. For comparison, the support was also input to Gao and Reeves' method [GR00]. In both selections, the number of samples measured was set to $p = q + 500$, corresponding to a reduction in sampling density of 28.6%. Running MATLAB® on a 1.7GHz AMD™ Athlon PC with 512MB of RAM, the selections took 497 seconds (8 minutes) and 29,075 seconds (8 hours), respectively. The CSS algorithm therefore gave a speed advantage of approximately 50 times over Gao and Reeves' method.

The measurements had noise added to the equivalent of 25 dB SNR and reconstructions were performed using Walsh and Nielsen-Delaney's method [WN94] (see Section 4.3). The sampling patterns selected are shown in (a) and (b) of Fig. 6.5 while the reconstructed images are shown in (c) and (d). The results for the CSS algorithm are shown on the left and for Gao and Reeves' method on the right. Numerical results are presented in Table 6.3, with RMS errors computed with respect to the true mathematical object. The ratio of trace metrics resulting from the two sampling schemes is 1.3 dB (i.e. with the sampling pattern produced by Gao and Reeves' method as a benchmark, $\tilde{\eta} = 1.3$ dB). Since the RMS errors were computed with respect to the true object, the majority of the error is due to truncation

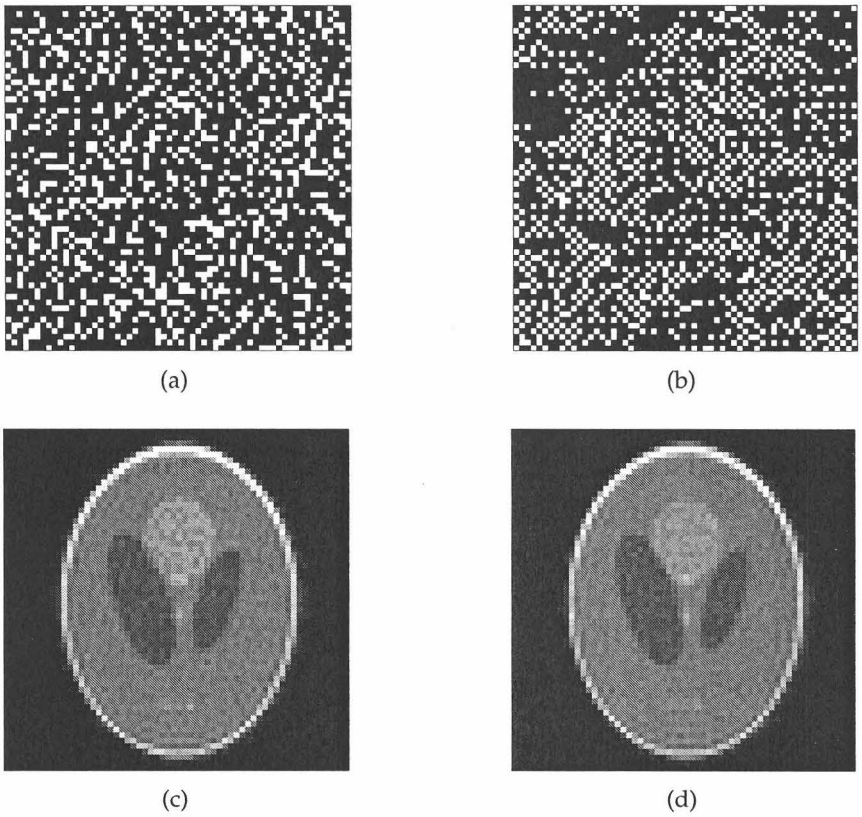


Figure 6.5 Results of the CSS algorithm for the Shepp-Logan head phantom, where the samples were corrupted by noise to the extent of 25 dB SNR: sampling patterns (black=measured) selected by (a) the CSS algorithm, and (b) Gao and Reeves' method; reconstructions from (c) sampling pattern (a), and (d) sampling pattern (b).

of k-space rather than to noise. This explains why the RMS errors are almost identical: the low value of $\tilde{\eta}$ predicts a small difference in RMS error anyway, but the truncation artefact dominates the error and suppresses the difference due to noise. Inspection of Fig. 6.5(c) and (d) shows that the two reconstructions are also comparable visually.

6.5 Periodic Cartesian sample selection

Having dealt with the difficult problem of Cartesian sample selection without periodicity, attention is now turned to the case of periodic Cartesian sampling, as developed in Chapter 5. The differences between PCS and CS were expounded in Section 5.5, the most significant being that the decomposition into subproblems effectively means a single sampling set is used with a number of different support sets. In Section 6.2, the 'tailor-made' sampling pattern approach of Reeves *et al.* was contrasted with the universal patterns approach of Bresler *et al.*—the latter is investigated here. However, instead of searching for universal patterns based on linear algebraic metrics, the focus is on developing easy-to-

Table 6.3 Numerical results for the reconstructions in Fig. 6.5

selection method	computation time (s)	RMS error	tmetric(·)
CSS	497	0.1037	6,836
Gao & Reeves	29,075	0.1009	5,029

compute heuristic metrics. The reader is referred to Section 4.4.7, where some definitions and properties of universal patterns were developed.

Once again, the mathematics presented is almost exclusively one-dimensional, with Theorem 4.2 used to apply the theory to higher dimensions. However, because the 1-D problem is representing the repeated block in periodic Cartesian sampling, only the elements of the size of the block are required to be pairwise relatively prime as opposed to the overall size as was the case for CS selection. For example, a 240×240 Cartesian grid could be sampled with a block size of (say) 5×12 or 15×8 under this scheme.

A useful universal pattern must allow accurate image reconstruction in the presence of noise over a wide variety of support sets. The contiguous pattern is universal (see Section 4.4.7) but in many cases exhibits bad conditioning. Consequently, a sampling pattern is to be found that (a) is universal, and (b) results in well-conditioned systems. Since universality is defined for the case $p = q$, this situation is often considered in the following subsections; this represents the worst-case scenario, since a universal pattern that produces well-conditioned systems will perform at least as well on problems with $q < p$.

In a similar methodology to selection for CS, two metrics are developed with the distinct roles of ensuring (a) universality (i.e. non-singularity), and (b) well-conditioned systems. Both metrics are based on the distances between samples, which were shown in Section 6.3.1 to be of fundamental importance for sampling systems.

6.5.1 Factors for universality

By definition, a universal pattern $\alpha \in U_{p,N}$ produces a non-singular system with any support set $\beta \in Q_{p,N}$. An argument for the role of factors in producing singular systems was presented in Section 6.3.3 above, and the joint factor metric that was developed combined information on the distances between sample locations that are factors (or multiples of factors) of N in both the sampling set, α, and in the particular support set, β. In attempting to identify sampling patterns that produce non-singular systems with *any* support set, the following modified metric (factor metric) is suggested that only considers the distances between sample locations in α; a larger (less favourable) metric is given to those patterns that

have more samples separated by distances that are factors (or multiples of factors) of N:

$$\mu^F(\alpha) = \sum_{x \in \mathcal{F}} \sum_{i=1}^{\lfloor N/2x \rfloor} \Delta(\alpha, ix), \tag{6.12}$$

where again $\mathcal{F}$ is the set of all factors of N (excluding 1 and N).

6.5.2 Repeated distances for conditioning

The second criterion for the desired sampling pattern is that it results in well-conditioned systems over all support sets. The trace squared metric, which depends on the particular support set being considered, was used in Section 6.3.2 above. It would be possible to use the worst-case metric over all support sets, but this would be extremely time-consuming; it would be preferable to use general rules that apply for all support sets, eliminating the need to assess individual support sets at all. Empirical evidence suggests that universal patterns containing the same distance between samples a number of times generally produce worse-conditioned systems than those with a smaller number of repeated distances. Consider the following example:

Example 6.2. Let $M = 1$, $N = 13$, and $p = 4$. (Note that since N is prime, all $\alpha \in Q_{p,N}$ are universal by Theorem 4.12; this aids analysis by removing the effect of factors from the causes of ill-conditioning.) Consider the following two universal patterns and their associated $\binom{p}{2} = 6$ distances between sample locations:

α	$\Delta(\alpha,1)$	$\Delta(\alpha,2)$	$\Delta(\alpha,3)$	$\Delta(\alpha,4)$	$\Delta(\alpha,5)$	$\Delta(\alpha,6)$
$\{0,1,5,9\}$	1	0	0	3	2	0
$\{0,1,4,6\}$	1	1	1	1	1	1

The first pattern has a maximum of three repeated distances while the second pattern has a maximum of only one. The worst-case trace metrics for the two patterns are 9,495 and 164, respectively, while the mean trace metrics over all unique support sets are 533 and 64, respectively. The second pattern clearly results in much better conditioned systems. ▲

While the mechanism is not well understood, investigations seem to show a noticeable relationship between the number of repeated distances and the conditioning performance of the universal pattern. Therefore, the following metric (repeated distances metric) is proposed, where a larger metric is less favourable:

$$\mu^{RD}(\alpha) = \max_{1 \le \delta \le \lfloor \frac{N}{2} \rfloor} \Delta(\alpha, \delta). \tag{6.13}$$

6.5.3 Periodic Cartesian sample selection algorithm

In contrast to selection for Cartesian sampling, a periodic Cartesian sampling scheme comprises two components: a block size, and a sampling pattern for the block. The block size directly affects the maximum possible reduction in sampling density and is dependent on the geometry of the support. The choice of a suitable block size is discussed shortly. As already stated, the methodology proposed in this chapter for the sampling pattern for the block is to use a universal sampling pattern. To this end, an algorithm is first developed to find a 1-D universal sampling pattern for a particular problem size, N and p.

Universal pattern selection algorithm

Analogously to the algorithm developed in Section 6.3.4, an SFS algorithm is proposed using the metrics developed above. The factor metric is used first to identify universal patterns followed by the repeated distances metric to select a universal pattern that results in well-conditioned systems. In addition, a complementary problem approach may be used, where a universal pattern with $N-p$ samples is found and the result complemented.

The algorithm, referred to as UPS1 (universal pattern selection 1-D), proceeds in an almost identical fashion to the CSS1 algorithm and is described precisely by the pseudo-code in Fig. 6.6. First, if $p > N/2$ then p is changed to $(N-p)$ for the purposes of selection. The first sample is then selected at the origin and the remaining $(p-1)$ selections are made sequentially using the factor metric followed by the repeated distances metric. A final random choice is made if necessary, but as implemented in the examples for this thesis the random choice is in fact made arbitrarily to be the sample with the lowest index that hasn't already been selected.

Like the CSS1 algorithm, the sequential nature of the algorithm allows for very efficient calculation of the metrics, giving computational complexity $\mathcal{O}(p^2N)$. MATLAB® code for the algorithm appears in Appendix B.

Block size selection

The particular block size used determines the supports of the subproblems, and the subproblem with the largest number of unknowns determines the minimum number of samples required per block for full reconstruction; this constraint is described by Eq. (5.59). Note that setting p_0 to the lower bound does not mean that the overall sampling density is equal to the minimum density: only the subproblem(s) with the largest number of unknowns would be exactly determined while all others would be over-determined (unless all subproblems happen to have the same number of unknowns). Consequently, the overall problem would be over-determined and hence sampled above the minimum density.

Algorithm *Universal-Pattern-Selection-1-D*(N, p)
Input: The problem order, N; and the number of samples to place, p
Output: A universal sampling pattern, α

/* Use complementary approach if applicable */
flag $\leftarrow 0$
if $p > N/2$
  then $p \leftarrow (N - p)$
    flag $\leftarrow 1$
/* Initialise sampling set and choices set */
$\alpha \leftarrow \{0\}$
$S \leftarrow \{1, 2, \ldots, N-1\}$
repeat
  $T \leftarrow S$
  /* Reduce choices using $\mu^F(\alpha)$ */
  $T \leftarrow \underset{e \in T}{\text{argmin}} \left[\mu^F(\alpha \cup \{e\})\right]$
  /* Reduce choices using $\mu^{RD}(\alpha)$ */
  $T \leftarrow \underset{e \in T}{\text{argmin}} \left[\mu^{RD}(\alpha \cup \{e\})\right]$
  /* Make a final random choice */
  $e \leftarrow$a randomly selected element of T
  /* Update sets */
  $\alpha \leftarrow \alpha \cup \{e\}$
  $S \leftarrow S \setminus \{e\}$
until $|\alpha| = p$
/* Compute complement if necessary */
if flag $= 1$
  then $\alpha \leftarrow \bar{\alpha}$
return α

Figure 6.6 Pseudo-code for the UPS1 algorithm. The argmin operator returns the set of arguments that produce the function's minimum value.

The sampling density can be made arbitrarily close to the minimum density by picking a sufficiently large block size [VB98], but a very large block size sacrifices much of the computational benefit of periodic Cartesian sampling. The discussion in Section 5.5 concluded that a moderately small block size (i.e. one with a total number of elements between, say, 20 and 100) provides a good tradeoff between complexity and reduction in sampling density.

It is proposed that several block sizes be initially tested and the one that gives the biggest reduction in sampling density be chosen. To ensure that a 'moderately small' block size is obtained, the block sizes tested should be only those that have a total number of elements between certain limits, say 20 and 100.

Overall selection algorithm

The overall selection algorithm simply comprises each of the previous steps in turn: a suitable block size is first selected and then a universal pattern obtained for the given parameters. Because a universal pattern is being found for a moderately small block size, the parameters p and N of the UPS1 algorithm are relatively small and consequently computation is very fast. Instead of finding a universal pattern with the minimum required value of p_0, it may be desirable to select a slightly higher value to yield better conditioned subproblems. The final step is to convert the 1-D universal pattern into an M-D sampling set. This algorithm is referred to as PCSS (periodic Cartesian sample selection).

6.6 Results for periodic Cartesian sample selection

Analogously to the ratio η introduced earlier, it is useful to define a ratio ζ to give the relative conditioning of a universal pattern:

$$\zeta(\alpha) = \frac{\max\limits_{\substack{\beta \in Q_{p,N} \\ \det(W_N[\alpha|\beta]) \neq 0}} \operatorname{tmetric}(W_N[\alpha|\beta])}{\min\limits_{\alpha' \in U_{p,N}} \max\limits_{\beta \in Q_{p,N}} \operatorname{tmetric}(W_N[\alpha'|\beta])}. \tag{6.14}$$

That is, $\zeta(\alpha)$, or simply ζ, gives the ratio of the worst-case trace metric of α to the worst-case trace metric of the best universal pattern (i.e. the one with the smallest worst-case trace metric). The ratio ζ (usually expressed in decibels) describes the increased level of noise that results from using the given sampling set with its worst-case support set rather than the best universal pattern with its worst-case support set. In order for ζ to still be meaningful if α is not universal, the numerator is defined to give the worst-case trace metric among the non-singular systems only. If α is universal, then $\zeta \geq 0\,\mathrm{dB}$ but otherwise $\zeta < 0\,\mathrm{dB}$ is possible. The universality of α can be measured by the proportion of unique

support sets for which a non-singular system results,

$$\rho = \frac{\left|\left\{\beta \;\middle|\; \beta \in \hat{Q}_{p,N},\ \det(W_N[\alpha|\beta]) \neq 0\right\}\right|}{\left|\hat{Q}_{p,N}\right|}, \tag{6.15}$$

where $\rho = 1$ indicates that α is universal.

In the following subsections, the two metrics are evaluated for small problem sizes and then the UPS1 algorithm is assessed on both small and large problem sizes. Finally, the overall PCSS algorithm is applied to 2-D and 3-D examples.

6.6.1 Performance of the metrics

Factor metric

To assess $\mu^F(\alpha)$ the cases for N prime were excluded since all values of $\mu^F(\alpha)$ would be zero and by Theorem 4.12 all such patterns are universal. For each of the 136 remaining small problem sizes, the subset of $\hat{Q}_{p,N}$ that is universal was determined together with the subset that had the minimum factor metric. The results are shown as a Venn diagram in Fig. 6.7, where a number in a particular region indicates the fraction of elements of $\hat{Q}_{p,N}$ (over the 136 problem sizes) that belong to that region. Fig. 6.7 shows that *every* universal pattern had the minimum factor metric, and *most* patterns (76%) with the minimum factor metric were universal. Furthermore those with the minimum factor metric that were not universal had a mean value of ρ of 0.975. Patterns with high values of ρ are 'almost universal' in that very few support sets actually result in singular systems. Since the application for universal patterns that is being considered involves a relatively small number of sup-

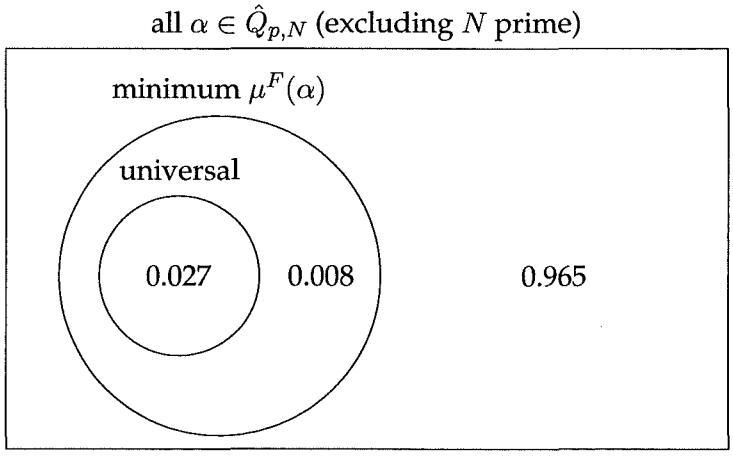


Figure 6.7 Venn diagram showing the success of $\mu^F(\alpha)$. The numbers indicate fractions of $\hat{Q}_{p,N}$ for small problem sizes (excluding N prime).

port sets (compared with the total number possible), it is highly unlikely that these 'rogue' support sets would be encountered by an almost universal pattern. Overall, these results indicate very good performance of the factor metric.

Repeated distances metric

All 190 small problem sizes were considered to evaluate $\mu^{RD}(\alpha)$. However, since the repeated distances metric is designed to identify those patterns that lead to well-conditioned systems from among all universal patterns, for each problem size the elements of $\hat{U}_{p,N}$ were considered instead of the elements of $\hat{Q}_{p,N}$. A Venn diagram of the results is shown in Fig. 6.8, with numbers indicating fractions of the elements of $\hat{U}_{p,N}$ (over small problem sizes). It is evident that 29% of the patterns with the minimum repeated distances metric exhibited $\zeta < 6\,\text{dB}$ conditioning, and a similar analysis shows that 47% give $\zeta < 10\,\text{dB}$ conditioning. Recall that ζ describes the difference in worst-case conditioning; the average difference in conditioning is significantly less. However, ideally these percentages would be higher and there are a large number of $\zeta < 6\,\text{dB}$-conditioned patterns that are not detected by $\mu^{RD}(\alpha)$, meaning the metric is not as effective as desired although a better alternative has been elusive. For example, any notion that the spread of the set has an effect can be discounted, since a consequence of Theorem 4.13 is that the contiguous pattern (which generally results in badly conditioned systems) has exactly the same best-case and worst-case trace metrics as other universal patterns that are more spread out.

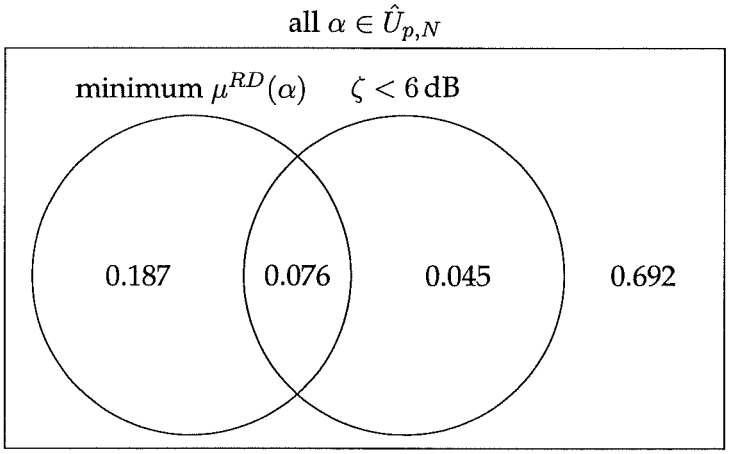


Figure 6.8 Venn diagram showing the success of $\mu^{RD}(\alpha)$. The numbers indicate fractions of $\hat{U}_{p,N}$ for small problem sizes.

Table 6.4 Results of the UPS1 algorithm for small 1-D problems

	range of ζ (dB)				total
	< 3	3–6	6–10	≥ 10	
universal	98	16	14	36	164
not universal	20	2	2	2	26
total	118	18	16	38	190

6.6.2 Performance of the universal pattern selection algorithm

Small one-dimensional problems

The UPS1 algorithm was applied to each of the 190 small problem sizes and a universal pattern was correctly found for 164 of these. For the remaining 26 problem sizes, the average value of ρ was 0.967 indicating that these patterns are almost universal. The results are shown in Table 6.4, with patterns classified by universality and ζ.

For the 164 universal patterns found here, 92 had $\zeta = 0$ dB, but in 48 cases this was obtained trivially since all universal patterns had the same worst-case trace metric. Interestingly, 16 of the 26 almost universal patterns had $\zeta < 0$ dB, which seems to suggest that almost universal patterns generally exhibit better performance than universal patterns (aside from the singular cases).

Large one-dimensional problems

For larger problem sizes, it is impractical to either exhaustively search for universal patterns or confirm whether a particular pattern is universal. However, to evaluate the algorithm's performance for larger sizes, 1,000 sizes were randomly selected such that $23 \leq N \leq 100$ and $2 \leq q \leq N - 2$. In each case, the UPS1 algorithm was run three times, for $p = q$, $p = q + 5$ (if applicable), and $p = q + 10$ (if applicable), in order to judge the performance with different levels of oversampling. The patterns produced by the algorithm were then tested against $1,000$ randomly chosen support sets.

The problem sizes selected resulted in 1,000 sizes for $p = q$, 904 sizes for $p = q + 5$, and 793 sizes for $p = q + 10$. In 989 of the problem sizes for $p = q$ (98.9%), a singular system was never found, indicating that the corresponding sampling patterns are likely to be universal. For the remaining 11 problem sizes for $p = q$ (1.1%), on average only 1.6 support sets (0.16% of the support sets tested) resulted in singular systems, which gives a mean estimated value for ρ of 0.998 and indicates that these patterns are probably almost universal. Singular systems were never found for $p = q+5$ or $p = q+10$. To measure conditioning, trace metrics

Table 6.5 Results of the UPS1 algorithm for large 1-D problems

oversampling[a]	'universal'[b]	not universal[c]	mean normalised tmetric(·) (dB)[d]
0	989	11	24.8
5	904	0	3.4
10	793	0	0.7

[a] number of samples added above the minimum density
[b] number of problem sizes for which a singular system was never found (i.e. the sampling pattern is probably universal)
[c] number of problem sizes for which a singular system was found
[d] mean value of the normalised trace metric in decibels (for the 'universal' cases)

were computed normalised by N for fair comparison (see Section 4.4.1). The mean values for the normalised trace metrics for the three levels of oversampling (in the universal cases) were 24.8 dB, 3.4 dB, and 0.7 dB, respectively. The results are shown in Table 6.5.

Overall, this indicates very good performance of the algorithm. Since it is often desirable to sample slightly above the minimum density (i.e. $p > q$), the excellent performance for $p = q + 5$ and $p = q + 10$ is noteworthy.

Two-dimensional example

To illustrate application of the UPS1 algorithm for $M = 2$, consider a 240×240 version of the Shepp-Logan head phantom with a relatively loose support (as used in Section 5.4 and shown in Fig. 5.8 on page 97). The performance of the UPS1 algorithm is demonstrated by comparing the sampling patterns produced with those computed via Gao and Reeves' method [GR01]. The results for two block sizes are presented here: analysis of the support showed that a 4×5 block with $p_0 = 13$ would yield full reconstruction with a 35% reduction in sampling density, while a 4×15 block with $p_0 = 38$ would provide full reconstruction with a 36.7% reduction. For the smaller size $p_0 = 13$ was chosen, while for the larger size $p_0 = 40$ was chosen; although this has decreased the reduction in sampling density to 33.3%, it has already been shown that slightly over-determined systems can provide much better conditioning than exactly determined systems, especially for comparatively larger problems (see Sections 4.5.3 and 5.5). The same block sizes and numbers of samples measured per block were used for Gao and Reeves' method.

The UPS1 algorithm was applied twice with parameters $N = 20$, $p = 13$, and $N = 60$, $p = 40$. For the two sampling patterns, the UPS1 algorithm took 0.01 and 0.02 seconds, respectively, running MATLAB® on a 1.7GHz AMD™ Athlon PC with 512MB of RAM.

Table 6.6 Numerical results for the reconstructions in Fig. 6.9

block size	selection method	computation time (s)	RMS error	sum of tmetric(·)
4×5	UPS1	0.01	0.087	11.81×10^4
	Gao & Reeves	94.7	0.079	8.92×10^4
4×15	UPS1	0.02	0.085	11.02×10^4
	Gao & Reeves	387.7	0.075	7.48×10^4

The 1-D patterns were then converted to 2-D using Theorem 4.2. On the same machine, an implementation of Gao and Reeves' method took 95 seconds (1.5 minutes) and 388 seconds (6.5 minutes) for the two sampling patterns, respectively, despite attempts to code the method as efficiently as possible. The UPS1 algorithm therefore gave a speed advantage of approximately 10^4 over Gao and Reeves' method.

Noise was added to the measurements to the equivalent of 20 dB SNR, and reconstruction performed using the PCSR algorithm developed in Section 5.3. The results are shown in Fig. 6.9: (a)–(d) show the central portion of the frequency domain; the UPS1 algorithm produced the blocks shown in (a) and (c), and Gao and Reeves' method those in (b) and (d). Reconstructions from these sampling patterns are shown in (g), (i), (h) and (j), respectively. For comparison, reconstructions from the full Nyquist set are shown in (e) and (f), where the measurements for (e) had noise added to the equivalent of 20 dB SNR, and the measurements for (f) had noise added to the equivalent of 18.2 dB SNR; the latter reflects the situation where an acquisition time 33.3% shorter is achieved by altering the MR imaging parameters. Numerical results are shown in Table 6.6. The ratios of sums of trace metrics resulting from the two selection techniques are 1.2 dB and 1.7 dB for the 4×5 and 4×15 block sizes, respectively.

In these examples, the sampling sets produced by the UPS1 algorithm gave reconstructions almost as good as those produced by Gao and Reeves' method, but the sampling patterns were selected in a fraction of the time. The first problem size ($N = 20$) is small enough that the universal pattern benchmarks can be used: the pattern produced by the UPS1 algorithm is universal and has the smallest worst-case trace metric (i.e. $\zeta = 0$ dB).

6.6.3 Performance of the overall algorithm

Three-dimensional example

The full PCSS algorithm is demonstrated here using the 3-D spin warp MRI data set normalised so that the grey levels cover the range $[0, 1]$ (see Section 3.8). Due to the pairwise

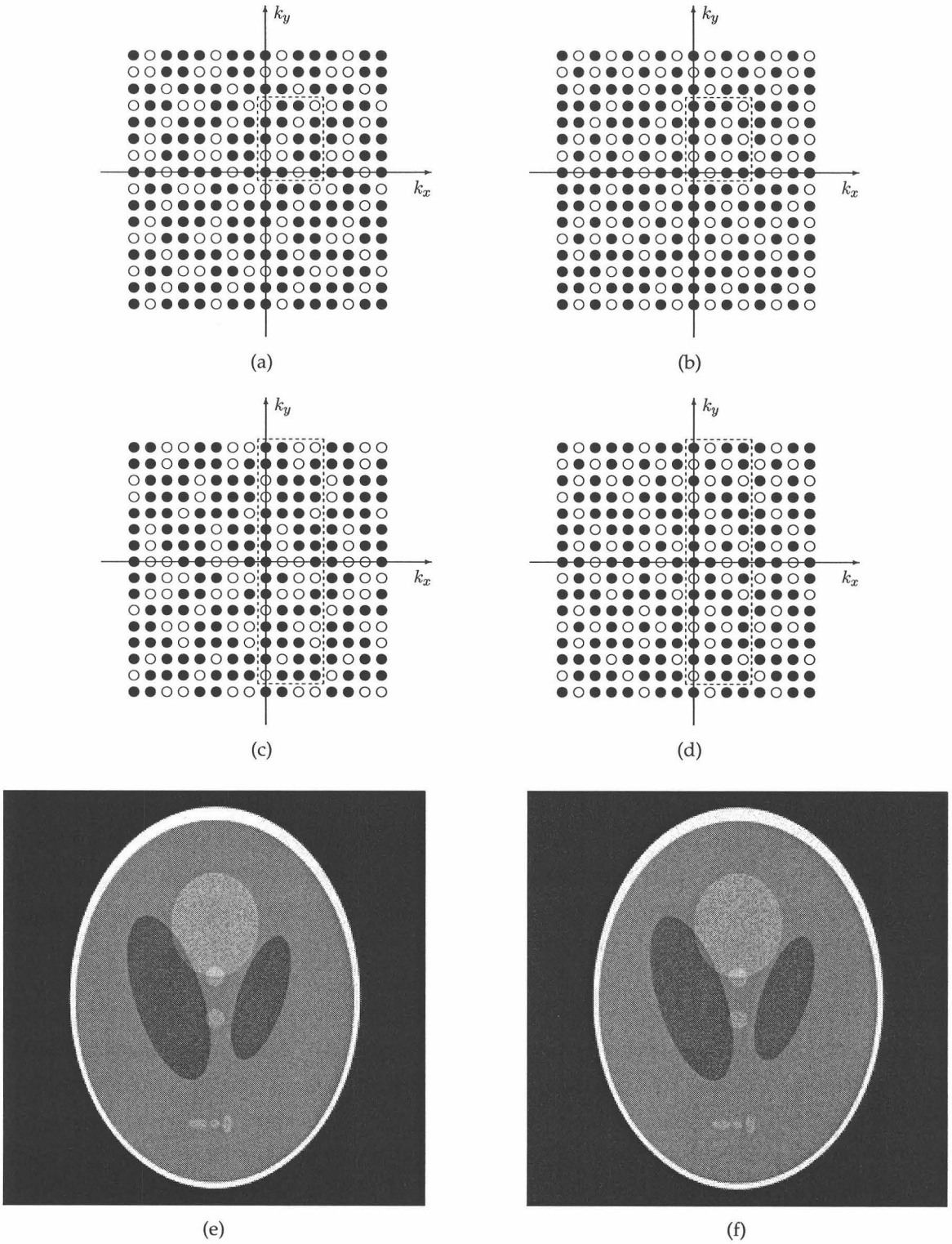
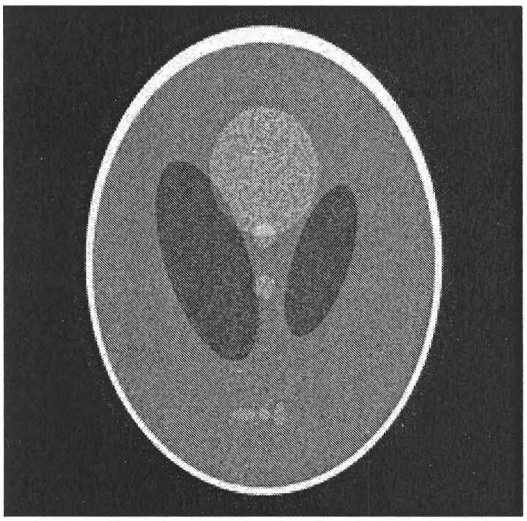


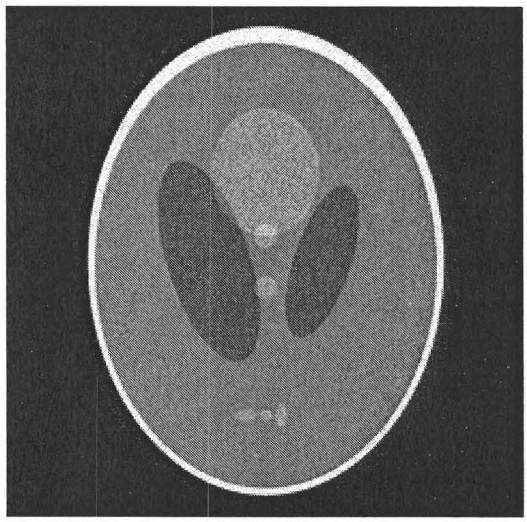
Figure 6.9 Results of the UPS1 algorithm for the Shepp-Logan head phantom where the samples were corrupted by noise to the extent of 20 dB SNR: (a)–(d) 16×16 central portions of k-space indicating the sampling patterns selected by (a) and (c) the UPS1 algorithm, and (b) and (d) Gao and Reeves' method; (e)–(f) reconstructions from the full Nyquist set for (e) 20 dB SNR, and (f) 18.2 dB SNR.

(g)

(h)



(i)



(j)

Figure 6.9 (continued from previous page) Results of the UPS1 algorithm for the Shepp-Logan head phantom where the samples were corrupted by noise to the extent of 20 dB SNR: (g)–(j) reconstructions from patterns (a), (b), (c) and (d), respectively.

relatively prime requirement for elements of the block size, the data set was truncated to $240 \times 240 \times 128$ since the number 240 has a large number of factors, many pairs of which are relatively prime. While the readout direction for this data is the anterior-posterior direction, for the purposes of this example it is assumed to be the superior-inferior direction, which means that a single point in the 2-D axial plane of k-space is measured per echo. The relevant 2-D support is determined by the region the object occupies in the x-y plane, and it is therefore analogous to a single axial slice as used in other examples in the thesis. More importantly, the support is likely to occupy a smaller proportion of the FOV than for an alternative readout direction. To determine the overall support, several individual slices were obtained from the IDFT of the full Nyquist set (mimicking the situation of several scout images). The supports for these images were determined and a binary OR operation performed to yield a single 'composite' support.

All results were performed using MATLAB® on a 1.7GHz AMD™ Athlon PC with 512MB of RAM. The PCSS algorithm was applied to the 240×240 composite support, with Gao and Reeves' method also used for comparison assuming a block size of 8×8. Note that this choice of block size was suggested by Gao and Reeves (see Section 6.2), but would be unsuitable for PCSS since the elements are not relatively prime. However, the first step of the PCSS algorithm is the selection a suitable block size: the limits to the total number of elements in the block were set to be 20 and 60, and block-oversampling by 2 was given as a parameter. The selection took 2.01 seconds, comprising 1.99 seconds for block size selection and only 0.02 seconds for the UPS1 algorithm. A block size of 4×15 was selected with 46 samples per block, giving a reduction in sampling density of 23.3%. For fair comparison, 49 samples per 8×8 block were used as parameters for Gao and Reeves' method to give a reduction in sampling density of 23.4%; the selection took 549 seconds (9 minutes).

Although the data is 3-D, the 2-D nature of the sampling and support sets gives the reconstruction problem considerable structure. It effectively implies periodicity with period one for both sets in the z-direction and as a result there are groups of 128 subproblems (i.e. the size of the z dimension) that share the same subsupport set. The system matrices for all 128 subproblems are hence identical and so the pseudo-inverse need be computed only once. The standard PCSR algorithm took 12 and 13 minutes for the sampling schemes selected by PCSS and Gao and Reeves' method, respectively, but when the 2-D structure was exploited in a modified PCSR algorithm, these times were reduced to only 42 and 44 seconds.

The results are shown in Fig. 6.10. The left-hand column shows the composite support and the sampling patterns selected by the PCSS algorithm and Gao and Reeves' method, respectively. The 'best estimate' was computed by IDFT of the full Nyquist set and three orthogonal slices (axial, coronal and sagittal) are shown in the second column. Slices from

Table 6.7 Numerical results for the reconstructions in Fig. 6.10

selection method	block size	samples per block	prop. reduc.[a]	computation time (s)	RMS error	sum of tmetric(·)
PCSS	4×15	46	0.233	2.0	0.0139	9.93×10^6
Gao & Reeves	8×8	49	0.234	548.6	0.0140	10.12×10^6

[a] proportional reduction in sampling density

the reconstructions are shown in the third and fourth columns, where the sampling patterns used were those selected by the PCSS algorithm for the third column and Gao and Reeves' method for the fourth column. The three slices shown from the reconstructions are the same three slices shown in the second column. All slice images are displayed on the range zero (black) to 0.3 (white) in order to enhance the contrast, meaning values above 0.3 are saturated.

The numerical results are shown in Table 6.7. The ratio of sums of trace metrics is -0.08 dB, which predicts that the reconstruction from the PCSS sampling pattern should have a slightly smaller error due to noise than the reconstruction from the sampling pattern due to Gao and Reeves; this is realised in a slightly smaller RMS error for the former. Inspection of Figs. 6.10(g)–(l) shows that the two reconstructions appear almost identical, and comparison with Figs. 6.10(b), (d) and (f) confirms that the reconstructions give good estimates of images from the full Nyquist set.

6.7 Conclusions

Selecting a Cartesian or periodic Cartesian sampling scheme is a difficult and computationally intensive problem, characterised by the enormous number of possible sample set combinations. Existing methods utilise linear algebraic techniques, resulting in very long computation times. The approach taken in this chapter has been to identify heuristics for non-singular and well-conditioned problems, with an emphasis on efficient computation. The result is a tradeoff between speed and conditioning, though sometimes no significant loss in conditioning performance occurs.

The CSS algorithm developed for Cartesian sampling provides a selection speed advantage of 50 times over an existing method in the example shown. For a moderately small level of oversampling, the decreased conditioning performance can be minimised. Further development of this approach could lead to a better understanding of the role of distances in the conditioning of sampling systems and eventually to even faster sample selection. An inherent difficulty in approaching this problem is that the exhaustive results which are vi-

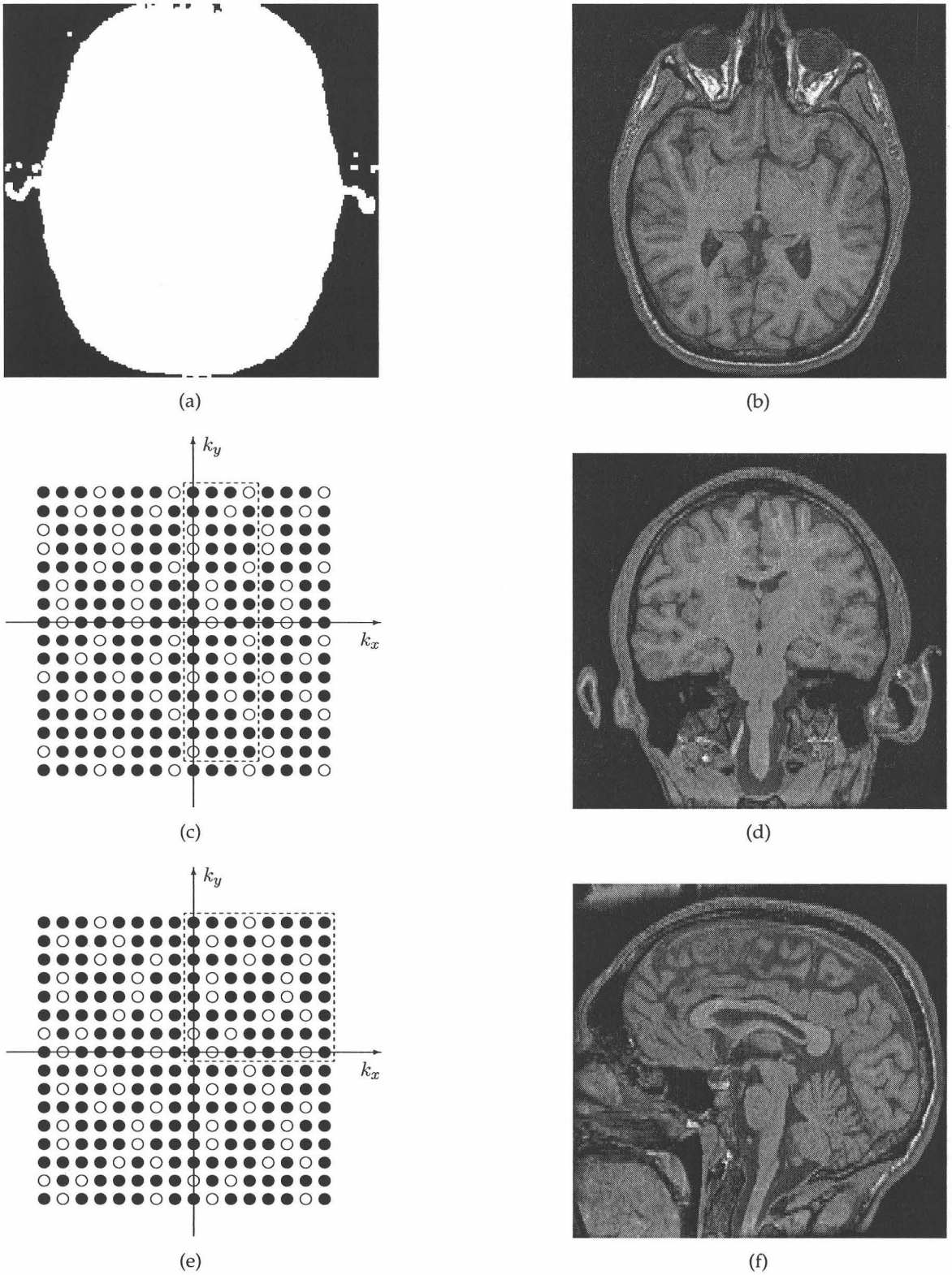
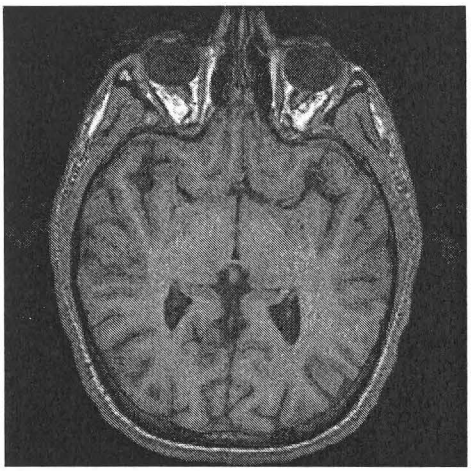
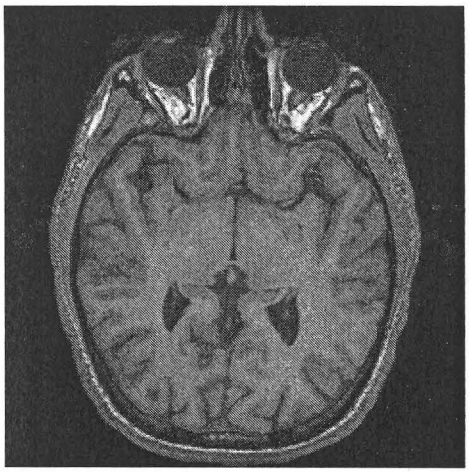


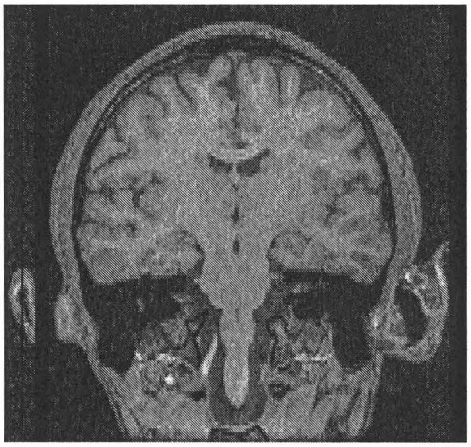
Figure 6.10 Results of the PCSS algorithm for 3-D MRI data: (a) support; 'best estimate' (IDFT of full Nyquist set) (b) axial, (d) coronal and (f) sagittal slices; 16×16 central portions of k-space indicating the sampling patterns selected by (c) the PCSS algorithm, and (e) Gao and Reeves' method.



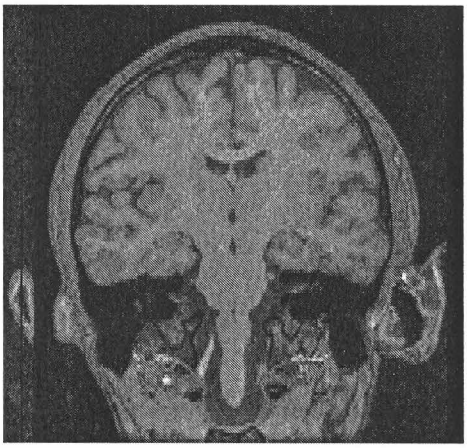
(g)



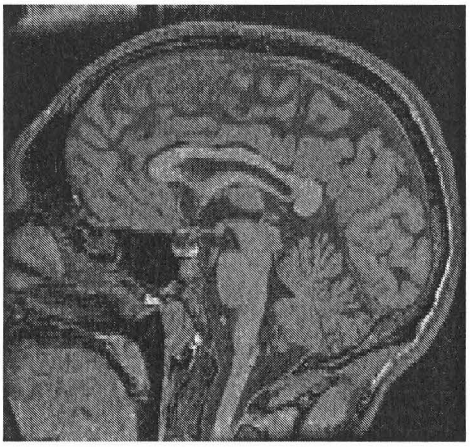
(h)



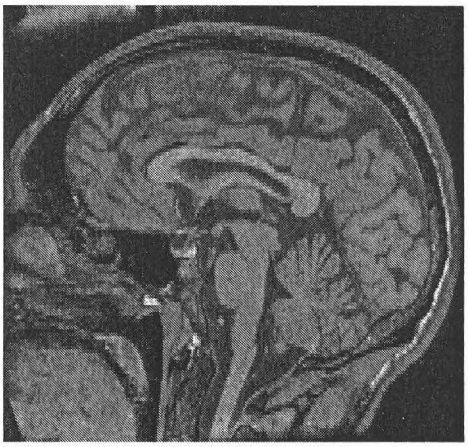
(i)



(j)



(k)



(l)

Figure 6.10 (continued from previous page) Results of the PCSS algorithm for 3-D MRI data: reconstructions from the sampling pattern selected by (g), (i), (k) the PCSS algorithm, and (h), (j), (l) Gao and Reeves' method.

tal for analysis are only available for small problem sizes; any characteristics observed for these sizes may not be representative of the sampling problem for larger sizes. This is more of an issue for CS rather than PCS, since the problem size is effectively greatly reduced in the case of periodicity.

The UPS1 algorithm provides almost instantaneous selection of sampling patterns—four orders of magnitude faster than an existing technique. The full PCSS algorithm took just two seconds in the example shown, but the majority of the computation time involved determining a suitable block size. It is possible that some block sizes may yield sizeable reductions in sampling density for a variety of different supports; the time required for block size selection could be drastically reduced by checking only a handful of possible block sizes. As an extension, the UPS1 algorithm could be used in conjunction with linear algebraic metrics such as those of Venkataramani and Bresler: a number of possible universal patterns could be established using the heuristic metrics, before more rigorous testing via linear algebraic techniques to make a final choice.

The techniques developed in this chapter shed light on the sampling problem, particularly with regard to the role of distances between sample locations in determining the performance of a sampling system. The selection algorithms developed provide an alternative approach to sample selection that allows fast, efficient computation—a unique feature among existing methods.

Chapter 7

Irregular sampling and reconstruction

7.1 Introduction

In many situations the sampling process may yield a set of irregularly spaced (or nonuniform) samples, either by practical necessity or by design. For example, in radio telescopy the locations of spectral samples are fixed by the physical siting of antennae and by the rotation of the earth. In MR imaging, on the other hand, it is possible to position samples in k-space virtually at will [LL00], allowing the flexibility to sample at irregular locations. As mentioned in Section 3.6.1, spiral sampling in MRI is an efficient way of traversing k-space in which samples are taken along a spiral trajectory commencing at the spectral origin [BSAG99]. Sampling along radial lines passing through the spectral origin is another irregular sampling scheme [GN93], used in cardiac imaging for example. The Cartesian and periodic Cartesian sampling schemes covered in Chapters 4 to 6 are also special cases of irregular sampling.

The fact that the sampling locations are irregularly spaced means that the FFT cannot be used and thus an alternative reconstruction technique is required. In radio astronomy, for example, the iterative CLEAN algorithm has been popular [Bro75], though the method relies to some extent on prior knowledge that stellar objects tend to be compact and isolated. The process known as 'gridding' has also been commonly used, in which the spectrum on a regular grid is estimated from the irregular set of measured samples [TB74]. This technique is also used in MRI [JMNM91] and is described in Section 7.2.

This chapter begins by describing some existing reconstruction techniques for irregular sampling. The problem is then investigated from the point of view of the Fredholm in-

tegral equation and a solution method derived. After some practical considerations, the method is applied to spiral sampling and Cartesian sampling sets. (Although the method is applicable to a general irregular sampling scheme, recall that Cartesian sampling is simply a special case of irregular sampling.)

7.1.1 Novel contributions

- **ISR algorithm** (Section 7.3)
  The reconstruction technique presented here is of limited novelty: a number of existing results are brought together and applied to the MRI problem. The technique is very similar to that in Refs. [VBM+00, DCA+02].

7.2 Existing methods

The 1-D irregular sampling problem was first addressed by Yen [Yen56] and his interpolation formula for the minimum-norm solution is given in Theorem 2.2. This formula has been subsequently derived in several different ways [CA87, CM00, Win92] and it can be expressed in M dimensions as follows: a bandlimited function $f(\mathbf{x})$ sampled at N locations, $\{\mathbf{x}_m\}_{m=0}^{N-1}$, has solution

$$\hat{f}(\mathbf{x}) = \sum_{m=0}^{N-1} b_m \phi(\mathbf{x} - \mathbf{x}_m), \tag{7.1}$$

where the b_m are functions of the sample values, $f(\mathbf{x}_m)$, and $\phi(\mathbf{x})$ is an interpolating function. The interpolator is chosen to be the inverse Fourier transform of the normalised indicator function of the bandregion (i.e. a function with value one inside the bandregion and zero outside) and the coefficients are given by $\mathbf{b} = \Phi^{-1}\mathbf{f}$, where Φ is the $N \times N$ matrix with elements $(\Phi)_{mn} = \phi(\mathbf{x}_m - \mathbf{x}_n)$ and $\mathbf{b}$ and $\mathbf{f}$ are vectors of the b_m and $f(\mathbf{x}_m)$, respectively.

In Ref. [CA87], the authors also derived three criteria for evaluating the performance of a sampling set based on the matrix Φ and this has been extended to consider the situation of different sized sets [Mil95]. Wingham investigated the problem in the presence of noise using the method of SVD [Win92] and showed that when the data is corrupted with additive, white, Gaussian noise, the least-squares estimator, the maximum likelihood estimator, and the maximum *a posteriori* probability estimator are all equal to the Miller-regularised Yen interpolation algorithm. Miller regularisation is a commonly used method of regularisation in which a small constant, ϵ, is added to the diagonal elements of the matrix prior to inversion.

Choi and Munson also investigated the Yen interpolator [CM98, CM00]. Unlike many studies that concentrate on deterministic formulations, in Ref. [CM00] the problem is in-

vestigated under a stochastic framework. When the data is corrupted with additive, white noise, the formula obtained corresponds to the Yen interpolator with Miller regularisation, where the magnitude of the regularisation parameter should be chosen according to the variance of the white noise. In MRI it is possible to estimate the noise level from areas of the image outside the body [Now99]. Even without noise, the roundoff error in numerical calculations may be regarded as white noise. If prior information in the form of the signal spectral density is known, then the autocovariance function of the signal process should be used as the interpolation kernel; for example, a Gaussian autocovariance models extreme smoothness and continuity.

The special case where the irregular samples are close to the Nyquist locations was also treated in Ref. [Yen56], with later work including an iterative procedure for uniform resampling [APSW89].

An alternative approach to the irregular sampling problem is the method of projection onto convex sets (POCS), which was first applied to image processing by Youla and Webb [You78, YW82]. Recall that the Gerchberg-Papoulis method described in Section 4.3 is in fact a special case of POCS [SS87]. Sauer and Allebach developed three algorithms that iterate between the spatial and spatial frequency domains [SA87]; the algorithms all apply the bandregion constraint in the spatial frequency domain but differ in their spatial domain constraints. This approach was extended by Yeh and Stark, who developed both an iterative and a one-step algorithm [YS90]. Interestingly, they note that the equation for their one-step method is a generalisation of the minimum-norm Yen solution discussed above. An overview of POCS applied to image recovery can be found in Ref. [SS87].

The mathematical literature also provides some results on irregular sampling, such as Refs. [Gro92, FG92]. The theory of frames, originating from the work of Duffin and Schaeffer [DS52], forms the background for many techniques including Ref. [San94] and the work by Benedetto *et al.* [BW99, BPW02], which addresses the problem of spiral sampling in MRI. An overview of frames and their relationship with irregular sampling is given in Ref. [Ben92]. After reformulating the reconstruction problem using 2-D trigonometric polynomials, Strohmer presents an iterative method for reconstruction [Str97].

Spiral sampling is becoming increasingly common in MRI [KFC95, BSAG99] due to its speed of data acquisition and the advent of faster reconstruction techniques. The gridding reconstruction algorithm [O'S85, JMNM91] operates by convolving the measured samples with a continuous window function, taking new samples on a Cartesian grid, and finally performing an IFFT. To compensate for the irregular sampling density, the original samples are first weighted using appropriate density compensation factors (DCFs); the choice of DCFs has been an active area of research [HKP97]. Instead of using the ideal infinite

sinc kernel for interpolation, the method employs a practical alternative such as a Kaiser-Bessel window. Using any kernel other than a sinc modifies the magnitude in the spatial domain, but this can be corrected by dividing by the point spread function of the window. This has the unwanted consequence of amplifying the aliasing error due to use of a finite window near the edges of the image; a common solution is to take new samples on a finer grid than necessary for the desired field of view, a technique called overgridding, and then to ignore the outer areas. An overview of gridding can be found in Ref. [WSB+01] and an analysis of the method's optimality is performed in Ref. [SN00]. The gridding algorithm can be made fast enough for real-time implementation through the use of lookup tables [DWD01] or fixed point processing [Lia00]. Other fast methods of reconstruction include the generalised FFT [Sar01], which is equivalent to gridding with a Gaussian kernel, and the nonuniform FFT [FS01].

While speed of reconstruction is often an important factor, in some imaging modalities the resulting SNR may be more important. For example, functional MRI usually involves relatively small images (64×64) but requires high SNR. In these circumstances, an alternative reconstruction algorithm is desirable. Van de Walle *et al.* developed a matrix-based reconstruction technique [VBM+00] that can be sped up by use of the conjugate gradient method [DCA+02].

A Bayesian approach to reconstruction was proposed by Marseille *et al.* [MFdB+94] [MdBF+96, GDMC+96], using the statistical prior that the histogram of intensity differences between adjacent pixels in an MR image can be modelled as a Lorentzian distribution [Fud89]. In Refs. [MFdB+94, MdBF+96], the authors investigated omitting phase encoding lines of k-space (k_y lines) and reconstructing using the Bayesian approach. They found that the choice of these lines is not very critical provided that limits are placed on the sizes of the gaps, the gaps at low k_y values are small, and the gaps are irregularly positioned [MdBF+96]. The choice of which phase encoding lines to omit has also been investigated under a statistical framework in Ref. [AB02]. Bayesian reconstruction has been further developed Wajer *et al.* [WdBF+98, WSB+01, WCL+02], including a means of accelerating the reconstruction (which can be time consuming) by weighting the likelihood term by the inverse of the sampling density [WCL+02].

The above discussion shows that the field of irregular sampling is alive with work from a variety of approaches. Partly because of this breadth, it was decided to investigate the problem from a standard starting point for inverse problems in signal processing—the Fredholm integral equation [PMBdM84].

7.3 Reconstruction technique

7.3.1 Theory

Consider the situation found in many imaging situations where it is necessary or convenient to measure the Fourier transform of an object rather than the object itself. The imaging problem may then be posed as the reconstruction of an M-D continuous object, $f(\mathbf{x})$, from a set of measurements taken in k-space, $F(\mathbf{k}_n)$, where $\{\mathbf{k}_n\}_{n=0}^{N-1}$ is a set of N locations in k-space. A Fourier transform relationship exists between $f(\mathbf{x})$ and $F(\mathbf{k}_n)$:

$$F(\mathbf{k}_n) = \int_{-\infty}^{\infty} f(\mathbf{x}) \exp(-i2\pi\mathbf{k}_n \cdot \mathbf{x})\, d\mathbf{x}, \tag{7.2}$$

or equivalently,

$$F(\mathbf{k}_n) = \int_{\mathcal{S}} f(\mathbf{x}) K(\mathbf{k}_n, \mathbf{x})\, d\mathbf{x}, \tag{7.3}$$

where $\mathcal{S}$ is the region of support and $K(\mathbf{k}, \mathbf{x}) = \exp(-i2\pi\mathbf{k} \cdot \mathbf{x})$.

Many problems in image reconstruction are characterised by a Fredholm equation of the first kind [PMBdM84] and inspection of Eq. (7.3) reveals that it is in fact a moment discretised version [Nas76, Rus87]. Define the operator $T_N : L_2(\mathcal{S}) \to \mathbb{C}^N$ by

$$T_N f = \begin{pmatrix} \langle f(\mathbf{x}), \bar{K}(\mathbf{k}_0, \mathbf{x})\rangle \\ \langle f(\mathbf{x}), \bar{K}(\mathbf{k}_1, \mathbf{x})\rangle \\ \vdots \\ \langle f(\mathbf{x}), \bar{K}(\mathbf{k}_{N-1}, \mathbf{x})\rangle \end{pmatrix}, \text{ and } \quad \mathbf{F} = \begin{pmatrix} F(\mathbf{k}_0) \\ F(\mathbf{k}_1) \\ \vdots \\ F(\mathbf{k}_{N-1}) \end{pmatrix}, \tag{7.4}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product in $L_2(\mathcal{S})$ and the bar denotes conjugation.

Eq. (7.3) may now be written in the form $T_N f = \mathbf{F}$ and the minimum-norm least-squares solution is given by [Nas76, Rus87]

$$\hat{f}(\mathbf{x}) = T_N^{+}\mathbf{F} = \sum_{n=0}^{N-1} b_n \bar{K}(\mathbf{k}_n, \mathbf{x}), \quad \mathbf{x} \in \mathcal{S}, \tag{7.5}$$

where T_N^{+} is the Moore-Penrose inverse of T_N and $\{b_n\}_{n=0}^{N-1}$ is a set of scalar coefficients (discussed below).

If $F(\mathbf{k})$ is a member of the range of the original infinite-dimensional operator, K, then $\hat{f}(\mathbf{x})$

can be written explicitly as

$$\hat{f}(\mathbf{x}) = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} F(\mathbf{k}_m)\gamma_{nm}\bar{K}(\mathbf{k}_n, \mathbf{x}), \quad \mathbf{x} \in \mathcal{S}, \tag{7.6}$$

where γ_{mn} are elements of the Moore-Penrose inverse of the $N \times N$ matrix A:

$$\begin{aligned} A_{mn} &= \langle K(\mathbf{k}_m, \mathbf{x}), K(\mathbf{k}_n, \mathbf{x}) \rangle \\ &= \int_{\mathcal{S}} K(\mathbf{k}_m, \mathbf{x})\bar{K}(\mathbf{k}_n, \mathbf{x})\, d\mathbf{x}. \end{aligned} \tag{7.7}$$

Furthermore, if the set $\{K(\mathbf{k}_n, \mathbf{x})\}_{n=0}^{N-1}$ is linearly independent then γ_{mn} may instead be defined as elements of the inverse of A and the set $\{b_n\}_{n=0}^{N-1}$ of Eq. (7.5) is uniquely determined by solving the matrix equation $A\mathbf{b} = \mathbf{F}$, where $\mathbf{b} = [b_0, b_1, \ldots, b_{N-1}]^T$.

Inspection of Eq. (7.6) shows that it has a form equivalent to the interpolation formula for the irregular sampling problem originally devised by Yen [Yen56] and discussed in the previous section. One main difference is that the solution is obtained directly in the spatial domain; although sampling is performed in the spatial frequency domain, there is no need for an intermediate step [Win92]. Before implementing Eq. (7.6), however, several practical issues need addressing.

7.3.2 Methods

The large number of measurements involved in a problem with a reasonably sized image results in an ill-conditioned algebraic system to solve [Rus87]. The presence of noise on the measurements exacerbates the problem. The use of Miller regularisation or pseudo-inversion via SVD was suggested in Ref. [Win92], while SVD or the conjugate gradient method [Lue84, She94b, GV96] were the methods employed in Refs. [VBM+00, DCA+02]. With SVD, the basis functions corresponding to the singular values that are too inaccurate to be useful (i.e. those closest to zero) may be left out altogether, while the iterative conjugate gradient method has inherent regularisation properties [BB98] that provide a more stable solution. Moreover, instead of an inverse or pseudo-inverse required to be found, matrix multiplication is the most computationally expensive operation and may be performed via a fast algorithm under certain conditions.

The derivation of the elements of matrix A in Eq. (7.7) is based on the object's region of support. A rectangular region is commonly used [CA87, VBM+00] resulting in a sinc function, but better results should be obtained from a support that better describes the object. As discussed in Section 3.8, it is envisaged that the support information could be gathered from preliminary scout images. There may be some benefit to using an analytical support

function (such as an ellipse) and it is straightforward to establish suitable ellipse parameters for a given support set [Wel91].

The specific reconstruction formulas for the MRI problem are

$$\hat{f}(\mathbf{x}) = \sum_{n=0}^{N-1} b_n \exp(i2\pi \mathbf{k}_n \cdot \mathbf{x}), \quad \mathbf{x} \in \mathcal{S} \tag{7.8}$$

$$A_{mn} = \int_{\mathcal{S}} \exp(-i2\pi(\mathbf{k}_m - \mathbf{k}_n) \cdot \mathbf{x})\, d\mathbf{x}. \tag{7.9}$$

The value of A_{mn} is simply the Fourier transform of the support region (i.e. a function with value one inside the support and zero outside) evaluated at location $(\mathbf{k}_m - \mathbf{k}_n)$. A convenient solution method is to first solve the matrix equation $A\mathbf{b} = \mathbf{F}$ and then use the result to apply Eq. (7.8). In Ref. [CM00], it was argued that if the statistics of the object are known they should be incorporated; adding a Lorenzian statistical prior into the kernel should provide a better reconstruction.

If the sampling set is a subset of the full Nyquist set (i.e. a Cartesian sampling set, as studied in Chapter 4), then matrix A has considerable structure—it is a submatrix of a block Toeplitz with Toeplitz blocks (BTTB) matrix. Strang proposed solving a Toeplitz system of linear equations by the conjugate gradient method, exploiting the fact that the matrix-vector multiplication step can be performed quickly via FFTs [Str86]. This technique is also applicable to a BTTB system, where the BTTB matrix is embedded in a BCCB matrix (see Section 2.3.3). Preconditioning is often used to speed up the conjugate gradient algorithm; a circulant preconditioner was proposed by Strang and many alternatives have been suggested by others, e.g. Refs. [Cha88, CO94]. Regularisation in the conjugate gradient algorithm is achieved through early termination of the iterations. Note, however, that it can be difficult to successfully combine the accelerated convergence due to preconditioning with early termination.

7.3.3 Algorithm

In light of the practical considerations discussed, the theory developed in Section 7.3.1 can now be implemented by the following algorithm:

1. Solve the matrix equation $A\mathbf{b} = \mathbf{F}$.
2. Form an image with any desired grid size using the coefficients, $\mathbf{b}$, obtained in step one and Eq. (7.8).

The matrix equation can be solved directly if the matrix is not too big (say less than 5,000 sample locations). Two methods of solution are investigated in the next section: matrix in-

version with Miller regularisation and the conjugate gradient method. If the sampling set is a Cartesian sampling set, the matrix can instead be used implicitly as part of the conjugate gradient method exploiting the BTTB structure. In either case, the conjugate gradient method should be terminated early to provide regularisation, possibly in conjunction with a preconditioner to accelerate convergence (the examples in this thesis do not use preconditioning but it will be shown that convergence is still rapid—the iteration is terminated if the residual error increases). Note that the coefficients actually describe a continuous image in the spatial domain and the image is discretised for display only. The algorithm is henceforth referred to as ISR (irregular sampling reconstruction). However, the particular method of solution is identified by a suffix: ISR+M or ISR+CG for Miller regularisation or the conjugate gradient method, respectively. For the special case of Cartesian sampling where the BTTB structure of the matrix can be exploited and the system solved using the conjugate gradient method, the algorithm is referred to as ISR+CG+FFT to emphasise that FFTs are used to perform the matrix multiplication step.

It should be noted that this technique is very similar to the work by Van de Walle *et al.* [VBM+00, DCA+02], although here use of a non-rectangular support is suggested and the special case of Cartesian sampling is considered. Both of these features are demonstrated in the next section.

7.4 Results

To demonstrate the ISR algorithm, two types of irregular sampling are considered: spiral and Cartesian sampling. Although other algorithms address reconstruction specifically for Cartesian sampling schemes, it is interesting to consider application of this more general technique to the special case.

7.4.1 Spiral sampling

In spiral MRI, the k-space trajectories are typically chosen as interleaved, outgoing, Archimedian spirals [BSAG99]. (The trajectories of multiple spirals with different starting angles at the origin of k-space fit between each other without touching, giving rise to the term interleaves.) The lth interleaf of a 2-D spiral centred at the origin with radius $k_{\max}$ and L interleaves is described by [BSAG99]

$$\begin{aligned} k_{x,l}(t) &= \Re\left\{k_{\max}\psi(t/T)\exp\left[i2\pi(R\psi(t/T)+l/L)\right]\right\} \\ k_{y,l}(t) &= \Im\left\{k_{\max}\psi(t/T)\exp\left[i2\pi(R\psi(t/T)+l/L)\right]\right\}, \end{aligned} \tag{7.10}$$

where T is the duration of the interleaf, R is the number of revolutions, and $\psi(t/T)$ is a monotonically increasing function with $\psi(0) = 0$ and $\psi(1) = 1$. A suitable choice for

Table 7.1 Numerical results for the reconstructions in Fig. 7.1

reconstruction method	computation time (s)	RMS error
gridding	124	0.1252
ISR+M	507	0.0952
ISR+CG	398	0.0971

$\psi(t/T)$ is

$$\psi(t/T) = \frac{t/T}{\sqrt{\alpha + (1-\alpha)t/T}} \qquad 0 \leq \alpha \leq 1. \tag{7.11}$$

The parameter α determines the type of spiral: $\alpha = 1$ gives a spiral with constant angular velocity, $\alpha = 0$ gives a spiral with constant linear velocity, and $0 < \alpha < 1$ gives a spiral with constant angular velocity for low t and constant linear velocity for high t. With $\alpha > 0$, the singularity at $t = 0$ for a constant linear velocity spiral can be avoided and the spiral optimised for k-space coverage subject to hardware constraints.

Consider the Shepp-Logan head phantom with a relatively loose support, as defined in Section 5.4. A single-interleaf spiral trajectory was chosen with parameters $k_{\max} = 32$ (corresponding to the extent of k-space covered by a 64×64 Cartesian grid), $R = 32$, and $\alpha = 0.3$, with sample locations obtained from 3,584 equally spaced values of t/T between 0 and 1. Samples were 'measured' at these locations then corrupted with pseudo-random, complex, white, Gaussian noise to the extent of 25 dB SNR.

The system matrix was explicitly created using the formula for the relatively loose elliptical support. The system of equations was then solved for the coefficients in two ways: using matrix inversion with Miller regularisation (with ϵ equal to the noise power), and by the (non-preconditioned) conjugate gradient method. Running MATLAB® on a 1.7GHz AMD™ Athlon PC with 512MB of RAM, the reconstructions took 507 and 398 seconds, respectively. The conjugate gradient method for ISR+CG terminated after nine iterations. For comparison, reconstruction was also performed by gridding using the Kaiser-Bessel window. All reconstructions were made onto a 256×256 grid and the results are shown in Fig. 7.1: (a) shows the spiral sampling set, (b) shows the true phantom (the dashed oval indicates the support that is assumed to be known *a priori*), (c) shows the reconstruction via gridding, and (d) and (e) show the reconstructions using ISR+M and ISR+CG, respectively. Numerical results are presented in Table 7.1, with RMS errors computed with respect to the true mathematical object (for the region inside the support only). Note that the gridding implementation used is not optimised for speed—a fast implementation would take less than a second.

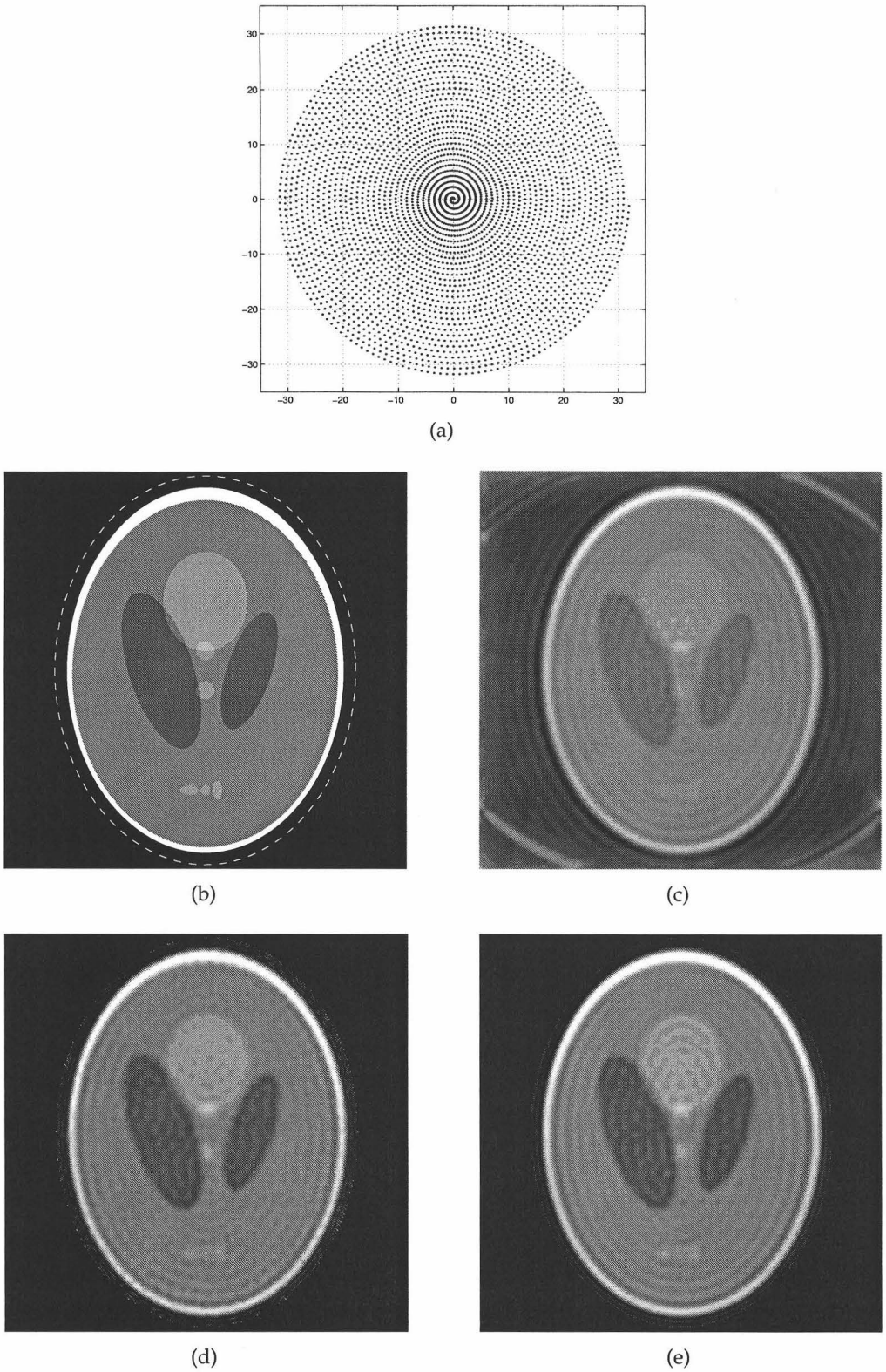


Figure 7.1 Results of the ISR algorithm for the Shepp-Logan head phantom: (a) spiral sampling pattern; (b) true phantom (the dashed oval indicates the support); reconstructions via (c) gridding, (d) ISR+M, and (e) ISR+CG.

Table 7.2 Numerical results for the reconstructions in Fig. 7.2

reconstruction method	computation time (s)	RMS error
G-P method	5	0.0964
ISR+CG+FFT	2,908	0.0954

The ISR reconstructions are comparable visually, numerically, and temporally. Although much faster, the gridding reconstruction is noticeably inferior in quality and this is confirmed numerically. This reconstruction technique therefore offers a tradeoff of reconstruction time for image quality, characteristics which are suitable for imaging situations such as fMRI, where relatively small sized sampling sets are used and SNR is a priority.

7.4.2 Cartesian sampling

Again consider the Shepp-Logan head phantom with the same support as above. Sampling locations were selected randomly from a full 128×128 Cartesian grid such that 30% of the samples were omitted. The measured samples were corrupted with pseudo-random, complex, white, Gaussian noise to the extent of 25 dB SNR. The size of the system matrix required for this sampling set makes the ISR+M and ISR+CG methods intractable given the computer equipment used. Instead, the BTTB structure of the system matrix can be exploited through the ISR+CG+FFT method.

Running MATLAB® on a 1.7GHz AMD™ Athlon PC with 512MB of RAM, the conjugate gradient method terminated after ten iterations and the entire reconstruction took 48 minutes. Interestingly, however, solving for the coefficients took only nine seconds! Since this would seem to be the more difficult task, it is possible that the second step of the algorithm (reconstructing onto a 256×256 grid) could be sped up. This type of sampling set is not suitable for reconstruction by either the PCSR algorithm developed in Chapter 5 (since the sampling is not periodic) or a direct matrix-based technique (due to the prohibitively large size), but for comparison, reconstruction was also performed using the iterative Gerchberg-Papoulis method (described in Section 4.3), which took five seconds on the same machine (the iterations were terminated once successive sum of square errors of the region outside the support were within 10^{-10}). The results are shown in Fig. 7.2: (a) shows the Cartesian sampling set (black=measured), (b) shows the true phantom, and (c) and (d) show reconstructions using the G-P method and ISR+CG+FFT, respectively. Numerical results are presented in Table 7.2, with RMS errors computed with respect to the true mathematical object. The reconstructions are comparable both visually and numerically, although this should not be too surprising since both methods have the same prior information and are related to Yen's minimum-norm solution.

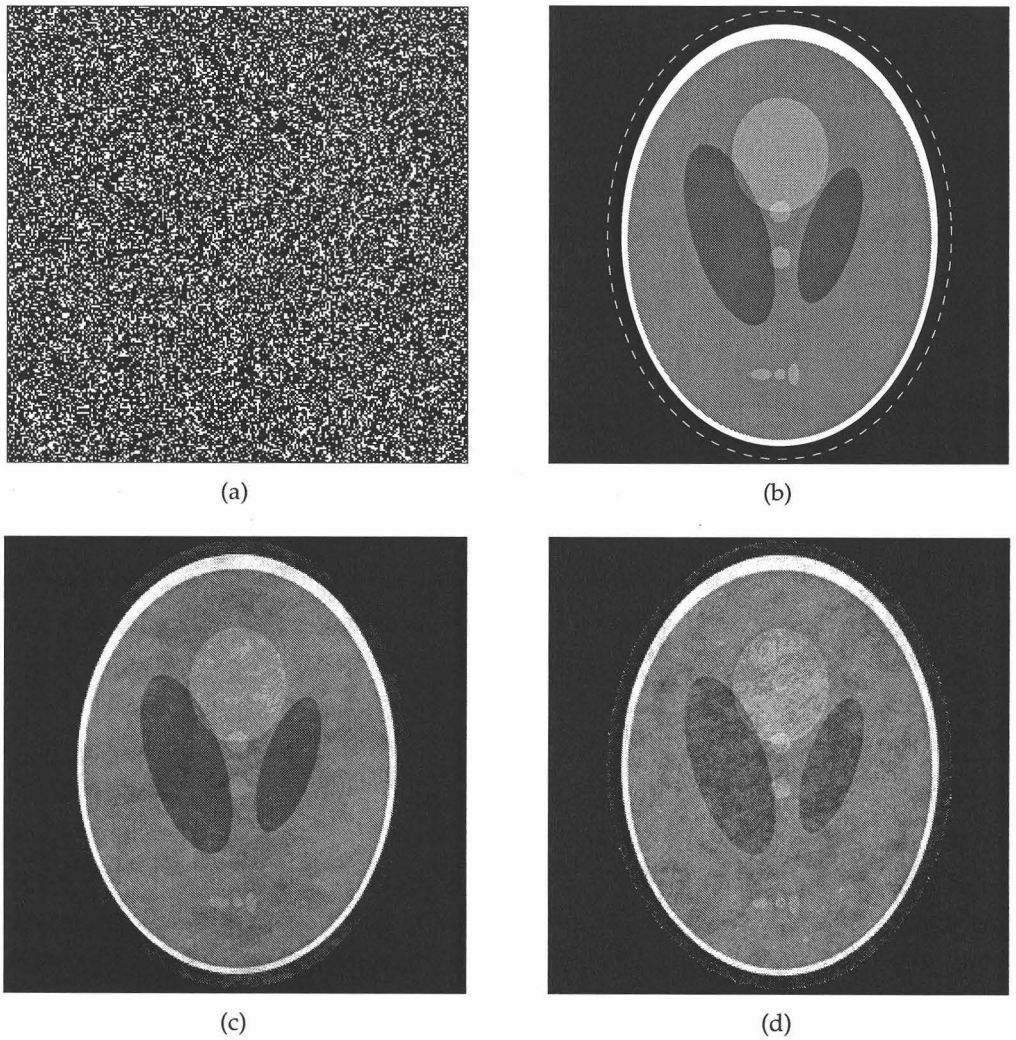


Figure 7.2 Results of the ISR algorithm for the Shepp-Logan head phantom: (a) Cartesian sampling pattern (black=measured); (b) true phantom (the dashed oval indicates the support); reconstructions via (c) Gerchberg-Papoulis method and (d) ISR+CG+FFT algorithm.

7.5 Conclusions

Reconstruction of a signal from a set of irregularly spaced samples is a non-trivial problem that has attracted solutions from a variety of approaches. It has been investigated here from the point of view of moment discretisation of the Fredholm equation of the first kind, and the reconstruction formula obtained has the same form as a standard irregular sampling technique due originally to Yen.

In practice, difficulties may arise in performing the numerical calculations, both in terms of stability of the solution and tractability. Two forms of regularisation have been considered and shown to give stable results in the presence of noise. For the special case of Cartesian sampling, the structure of the system matrix may be exploited to yield an FFT-based solution.

The techniques proposed require longer computation than alternatives such as gridding or the G-P method. Although there is not a significant improvement in reconstruction for the Cartesian sampling example, the ISR algorithm results in better reconstruction compared to gridding for spiral sampling, thereby offering a tradeoff between reconstruction time and image quality. The ISR algorithm is therefore suitable to applications in which SNR is a priority. The relatively small sized images for fMRI are well-suited to the matrix-based versions of the technique.

Chapter 8

Conclusions and future work

Many imaging systems are designed to estimate a spatial domain object with limited support from data sampled in the spatial frequency domain. In MRI, a desire to speed up the imaging process motivates the use of irregular sampling schemes, while in other systems irregularity may be imposed as a result of missing samples or by the nature of the imaging process. The selection of such a sampling scheme and the subsequent image reconstruction are the problems investigated in this thesis.

The term Cartesian sampling is used in this thesis to describe the case where samples are constrained to lie on a Cartesian grid, but not all grid points are sampled. This is applicable to MR applications in which a single point in 2-D k-space is measured per echo (such as MRSI and 3-D MRI) and to other imaging procedures in which such a data set is collected, perhaps due to samples being missing or corrupted. A mathematical framework for Cartesian sampling was introduced in Chapter 4 and the sampling problem posed. To facilitate analysis, the question of how to measure the conditioning of a matrix was considered and this led to use of the trace metric—an alternative to the conventional condition number defined by the reconstruction error when the measurement vector is corrupted with additive, white, Gaussian noise. A body of theory was then presented that has practical implications for how to approach the problem and provides intuition about its nature. Sampling issues concerning Hermitian symmetry, sampling spread and problem size were also discussed.

The computational complexity involved in the Cartesian sampling problem motivates the use of a sampling scheme that reduces this complexity. This is realised by the subject of Chapter 5, periodic Cartesian sampling, in which a small sampling pattern is repeated throughout a Cartesian sampling scheme. The key result is that periodicity in the sampling pattern results in sparsity in the point spread function. This fact was exploited in development of the periodic Cartesian sampling reconstruction (PCSR) algorithm in which

the overall problem is split into independent subproblems that can be solved separately. The same idea was used to develop the periodic Cartesian sampling prediction (PCSP) algorithm to predict the regions that could be reconstructed from a particular sampling scheme and support, and which could be used to evaluate candidate sampling sets before measurements are obtained. The link between subproblems and the overall problem was demonstrated mathematically: the singular values of the overall problem are distributed among the subproblems. The relationship between conditioning, reduction in sampling density, and size of the repeated sampling block was also investigated and it was shown that use of a periodic Cartesian sampling scheme can provide similar performance (i.e. the combination of conditioning and reduction in sampling density) to an oversampled ordinary Cartesian sampling scheme.

Both Cartesian and periodic Cartesian sampling involve measuring subsets of the full Nyquist set; the task of selecting a *good* subset was considered in Chapter 6. It was argued that the distances between sample locations is the crucial factor in determining the conditioning performance of a sampling set. Furthermore, it was conjectured that system singularity is closely related to when the distances between sample locations are factors of the problem size. Heuristic metrics were developed to incorporate these ideas and were used to form the basis of sequential forward selection algorithms for both Cartesian and periodic Cartesian sampling schemes (the Cartesian sample selection(CSS) and periodic Cartesian sample selection (PCSS) algorithms, respectively). To address the fact that in periodic Cartesian sampling, a single sampling set is effectively partnered with multiple support sets, the notion of universal patterns was adopted. This led to the universal pattern selection 1-D (UPS1) algorithm that underlies the PCSS algorithm; using this approach, a sampling pattern can be selected in the order of one second, certainly fast enough for practical situations.

Stepping back from the special case of Cartesian sampling, the general case of irregular sampling was considered in Chapter 7. After posing the problem from the point of view of a standard equation in inverse problems, a solution method was derived that was found to have the same form as the interpolation formula due originally to Yen. Prior information of the support was incorporated into the irregular sampling reconstruction (ISR) algorithm and practical issues of noise sensitivity were discussed. In applying the method to the case of Cartesian sampling, the system matrix was found to possess considerable structure (block Toeplitz with Toeplitz blocks). Exploiting this structure in the conjugate gradient algorithm using FFTs makes the method tractable even for large images.

At the outset, three questions were posed for the research:

1. How few samples can be measured while still enabling a good image to be formed?
2. Where should the samples be placed?
3. What reconstruction algorithm should be used?

Each of these is now considered in relation to the results established in the thesis.

8.1 Reduction in sampling density

The standard WKS sampling theorem implies a rectangular FOV in the spatial domain. When the object's support is smaller than the FOV, sampling theory (due to Landau) dictates that the minimum possible sampling density is given by the area of the support. This results in a linear relationship between the proportion of the FOV known *a priori* to be zero and the maximum possible proportional reduction in sampling density.

In the discrete environment, the particular shape of the support and the locations of the samples determine the conditioning of the resulting system of equations. The evidence in Section 4.5.3 suggests that solving the exactly determined (i.e. minimum density) problem becomes more difficult in some sense as the problem size increases. Specifically, for the particular support involved there may not exist a sampling pattern that results in an especially well conditioned problem and this becomes increasingly common for larger problem sizes. While it may be possible to approach the minimum density in the absence of noise, a noisy environment necessitates the use of oversampling to maintain an acceptably conditioned problem. (Note that this is *over*sampling with respect to the minimum density but *under*sampling with respect to the Nyquist density.)

The benefits of oversampling are highlighted by Gao and Reeves' figure (see Fig. 6.1), where oversampling by just a few samples can result in much better conditioning. However, the marginal gain from oversampling decreases as the level of oversampling increases (diminishing returns), as observed in Example 5.5. The general behaviour of oversampling providing better conditioning is supported by Theorem 4.1.

Although there are clear benefits to using periodic as opposed to ordinary Cartesian sampling from a computational point of view, there seems to be an inherent sacrifice in terms of the reduction in sampling density that can be obtained, as summarised by Eq. (5.59). This topic was investigated in Section 5.5 while taking care to also consider the effect on conditioning. In Example 5.5, it was demonstrated that some combinations of conditioning and reduction in sampling density can be obtained by either a Cartesian sampling scheme or a periodic Cartesian sampling scheme with a moderately small block size. In other words, the computational benefits of periodic Cartesian sampling need not necessarily require a tradeoff of conditioning and sampling density performance.

8.2 Sample placement

With such a large number of possible combinations of sample locations, determining sample placement is a difficult problem. A simplification was made in this thesis by considering sample selection only under Cartesian and periodic Cartesian sampling schemes.

The groundwork laid in Chapter 4 includes the identification of equivalent problems (Section 4.4.3), which simplifies analysis. The special case of a 1-D contiguous support was conjectured to require a maximally spread sampling set to produce an optimally conditioned system (Conjecture 4.9) and it was suggested that a similar result might exist in higher dimensions. However, investigations into spread in Section 4.5.2 did not reveal a general link between spread and conditioning; instead they raised questions about the importance of information content in evaluating the efficacy of a reconstruction. Although imperfect, the widespread use of RMS error is not without justification: it emphasises the importance of the general structure of an image and is easy to compute.

To cope with the large number of possible sampling combinations, sequential selection greedy-style algorithms were employed. The basis of these algorithms lies in heuristic metrics as opposed to linear algebraic metrics as found elsewhere in the literature, thereby providing a substantial speed advantage at the expense of a slightly sub-optimal sampling set. The speed of the PCSS algorithm in particular may make this tradeoff acceptable in some circumstances.

8.3 Image reconstruction

The same computational benefits that can be exploited in the selection of periodic Cartesian sampling schemes can be used in image reconstruction. This is demonstrated by the PCSR algorithm developed in Section 5.3. Furthermore, the predictive PCSP algorithm allows the determination of recoverable and unrecoverable regions before data is collected. The relatively small matrices involved allow calculation of conditioning information and the inherent parallelism in the algorithm could also be used to increase speed on a suitable computer architecture.

In contrast, the ISR algorithm developed in Section 7.3 provides a tradeoff of speed in favour of a better reconstruction. Use of support information improves the reconstruction and regularisation can be used to stabilise the solution. The speed and tractability of the algorithm can be improved in the case of Cartesian sampling by exploiting the resulting matrix structure.

8.4 General conclusions

In MR imaging, there exists a strong desire to reduce the data acquisition time. This inevitably results in a loss of SNR but in some applications this may be an acceptable tradeoff. The main method considered in this thesis for reducing acquisition time is to measure fewer samples than would ordinarily be measured; in MR modalities that measure a single point in 2-D k-space per echo, this provides a linearly proportional reduction in acquisition time. In other imaging procedures, the omission of samples may be unavoidable and it remains only to determine a suitable reconstruction technique and perhaps also the sample locations. To enable full reconstruction when sampling below the Nyquist density, prior knowledge must be available such as the object's support. In MRI this can be established relatively easily from scout scans.

The additional computational costs incurred from sampling only a subset of the full Nyquist set are threefold: the support must be found, a sampling scheme must be selected, and more complex image reconstruction must be performed. The remaining cost is image quality: apart from special cases, the combination of sampling below the Nyquist density and using prior knowledge of the support generally results in a reduction in SNR. The decision of whether or not to reduce the number of samples depends on the application: Is the loss in SNR acceptable? Is the additional computation time for selection and reconstruction tolerable? Further tradeoffs can be made between computation times for selection and/or reconstruction and SNR. As an additional variable, the use of a periodic Cartesian sampling scheme (if suitable) does not involve a large cost compared to ordinary Cartesian sampling but makes both selection and reconstruction simpler computationally.

As an illustration, consider the 3-D example in Section 6.6.3. This involved a 23.3% reduction in sampling density, which corresponds to a reduction in acquisition time from 12 minutes to just over nine. The support was obtained in three seconds, sample selection was completed in only two seconds, and the reconstruction took just 42 seconds. In this case, the reduced scan approach would result in a net decrease in imaging time of about two minutes. Importantly, however, the patient would require three minutes less in the scanner itself, meaning less chance of movement artefacts and the opportunity to begin other scans sooner. (Note that the next scan can begin regardless of whether post-processing of the previous scan has been performed.)

The use of an irregular sampling scheme such as spiral sampling certainly yields a reduction in acquisition time. The tradeoff investigated in this thesis was between reconstruction time and image quality. If it is not essential to view the images immediately, a longer reconstruction time that results in improved SNR may be desirable.

In the course of the research, the mathematical properties of the problem have been analysed and a variety of results obtained. It is hoped that these may also be useful in other areas and could provide the basis for further work in the field. The main objectives of the research were to investigate tools for sample selection and image reconstruction in the context of MRI—this thesis contains the development of novel and effective techniques for both selection and reconstruction that can be applied to MRI and other imaging situations.

8.5 Future work

One of the most important tools in image reconstruction is a basis for comparing different images. While RMS error has been used in this thesis, it would certainly be preferable to use a metric that better suits the end observer, usually human. For example, Ref. [SDWW99] uses a model of the human visual system to compare images.

A related problem is how to measure the sensitivity of a matrix system to noise, since this is intended to provide information on the accuracy of the solution. Although widely used, the 2-norm condition number only uses information from the maximum and minimum singular values. It was argued in this thesis that the metric $\mathrm{trace}(A^H A)^{-1}$ is superior, since it involves all singular values and is equivalent to the sum of square errors when the measurement vector is corrupted with additive, white, Gaussian noise. However, this is again related to the idea of RMS error and a better evaluation would presumably be obtained with more consideration of the human observer.

In Ref. [CA87], various properties of the system matrix were used to describe the 'goodness' of a sampling set. However, like RMS error they are based on *mean* error which does not always give a good description of the reconstruction; although the reconstruction may be very accurate in the vicinity of the sample points, it could be quite inaccurate in other places. The idea of assessing reconstruction at different locations was considered in Ref. [Tar97]. In the end, a combination of approaches is probably the best way to judge reconstruction.

One limitation of the selection methods developed in the thesis is the requirement that elements of the size of the discrete sequences be relatively prime. This arises from the equivalence between 1-D and M-D problems and a desire to simplify analysis by concentrating on 1-D. Higher-dimensional equivalents of relationships such as that between singularity and distances between sample locations would allow more widespread application of the universal pattern algorithm, for example. The obvious advantages of the FFT also motivate special consideration of the Fourier matrix of order 2^n. It would also be interesting to consider optimal preconditioners for the conjugate gradient algorithm in the special case of a submatrix of the Fourier matrix for M-D sequences.

Another interesting result observed in 1-D was that concerning a contiguous support and a maximally spread sampling set. Since so many real-world objects are contiguous (cross sections of the body, stars, etc.) an analogous result in 2-D would certainly be useful. However, the fact that there are such a large number of possible contiguous supports in dimensions greater than one means that there is almost certainly not a single optimal sampling set, although there may be some general rules that could allow 'pretty good' conditioning to be achieved the majority of the time. This type of analysis of the problem may also result in insight into the roles of distances between samples, repeated distances, spread of a sampling set, etc., in determining conditioning. In turn, this would allow the heuristic metrics introduced in the thesis to be fine-tuned to yield better conditioned problems.

The selection methods considered in this thesis were for Cartesian and periodic Cartesian sampling schemes, both of which constrain the possible sampling locations to a large degree, namely that they must lie on a Cartesian grid. Since there is no requirement in MRI that this be the case, the question that naturally arises is whether an improvement could be achieved by relaxing this constraint. One approach could be to simply use a progressively finer Cartesian grid, but this increases the complexity of what is already a difficult problem. In addition, intuition suggests that spacing samples very much closer than the Nyquist distance would not provide great gains. Approaching the problem without a Cartesian grid constraint at all is perhaps a better option, meaning that the problem is now the selection of an irregular sampling scheme. It is difficult to know the best optimisation approach; for example, it is likely that the optimal locations for (say) three and four samples would be considerably different, meaning that a sequential selection method would be unsuitable.

Finally, while not considered in this thesis, making use of statistical information in both selection [AB02] and reconstruction [WCL+02] is an important area of future work.

Appendix A

Lemmas for Section 5.3.4

Lemma A.1. *Let* $A^\alpha = W_{\mathbf{N}}^{-1}\,\mathrm{diag}(\mathbf{H}^\alpha)W_{\mathbf{N}}$, $A^\phi = W_{\mathbf{C}}^{-1}E_\phi^T E_\phi W_{\mathbf{C}}$, *and* $D_i = I[\gamma^i|\cdot]$. *Then,*

$$D_i A^\alpha D_j^T = \begin{cases} A^\phi & i = j \\ 0 & i \neq j. \end{cases} \tag{A.1}$$

Proof. First, note that the elements of A^α are

$$(A^\alpha)_{kl} = \frac{1}{\sqrt{N}}\, h^\alpha[\mathcal{S}^{-1}(k,\mathbf{N}) - \mathcal{S}^{-1}(l,\mathbf{N})], \tag{A.2}$$

the elements of A^ϕ are

$$(A^\phi)_{kl} = \frac{1}{\sqrt{C}}\, h^\phi[\mathcal{S}^{-1}(k,\mathbf{C}) - \mathcal{S}^{-1}(l,\mathbf{C})], \tag{A.3}$$

the periodicity in α means that

$$h^\alpha[\mathbf{n}] = \begin{cases} \sqrt{L}\, h^\phi[\mathbf{n} \circ^{-1} \mathbf{L}] & \mathbf{n} = \mathbf{m} \circ \mathbf{L},\ \mathbf{m} \in \mathbb{Z}^M \\ 0 & \text{elsewhere,} \end{cases} \tag{A.4}$$

and finally that

$$\gamma_k^i = \mathcal{S}^{-1}(i,\mathbf{L}) + \mathcal{S}^{-1}(k,\mathbf{C}) \circ \mathbf{L}. \tag{A.5}$$

Now, since $D_i A^\alpha D_j^T = A^\alpha[\gamma^i|\gamma^j]$, the elements of $D_i A^\alpha D_j^T$ are

$$\begin{aligned}(A^\alpha[\gamma^i|\gamma^j])_{kl} &= (A^\alpha)_{\gamma_k^i \gamma_l^j} \\ &= \frac{1}{\sqrt{N}} h^\alpha[\mathcal{S}^{-1}(\gamma_k^i, \mathbf{N}) - \mathcal{S}^{-1}(\gamma_l^j, \mathbf{N})] \\ &= \frac{1}{\sqrt{N}} h^\alpha[\mathcal{S}^{-1}(\mathcal{S}(\gamma_k^i, \mathbf{N}), \mathbf{N}) - \mathcal{S}^{-1}(\mathcal{S}(\gamma_l^j, \mathbf{N}), \mathbf{N})] \\ &= \frac{1}{\sqrt{N}} h^\alpha[\gamma_k^i - \gamma_l^j],\end{aligned}$$

and by Eq. (A.5),

$$\begin{aligned}&= \frac{1}{\sqrt{N}} h^\alpha[\mathcal{S}^{-1}(i, \mathbf{L}) + \mathcal{S}^{-1}(k, \mathbf{C}) \circ \mathbf{L} - \mathcal{S}^{-1}(j, \mathbf{L}) - \mathcal{S}^{-1}(l, \mathbf{C}) \circ \mathbf{L}] \\ &= \frac{1}{\sqrt{N}} h^\alpha[\{\mathcal{S}^{-1}(i, \mathbf{L}) - \mathcal{S}^{-1}(j, \mathbf{L})\} + \{\mathcal{S}^{-1}(k, \mathbf{C}) - \mathcal{S}^{-1}(l, \mathbf{C})\} \circ \mathbf{L}]. \qquad \text{(A.6)}\end{aligned}$$

Because α is periodic, $h^\alpha[\mathbf{n}]$ is only non-zero if $\mathbf{n} = \mathbf{m} \circ \mathbf{L}$, $\mathbf{m} \in \mathbb{Z}^M$. In Eq. (A.6), this can only occur if $\{\mathcal{S}^{-1}(i, \mathbf{L}) - \mathcal{S}^{-1}(j, \mathbf{L})\} = 0$, which is equivalent to the condition $i = j$. In this case,

$$\begin{aligned}(A^\alpha[\gamma^i|\gamma^j])_{kl}\Big|_{i=j} &= \frac{1}{\sqrt{N}} h^\alpha[\{\mathcal{S}^{-1}(k, \mathbf{C}) - \mathcal{S}^{-1}(l, \mathbf{C})\} \circ \mathbf{L}] \\ &= \frac{1}{\sqrt{N}} \sqrt{L}\, h^\phi[\mathcal{S}^{-1}(k, \mathbf{C}) - \mathcal{S}^{-1}(l, \mathbf{C})] \\ &= \frac{1}{\sqrt{C}} h^\phi[\mathcal{S}^{-1}(k, \mathbf{C}) - \mathcal{S}^{-1}(l, \mathbf{C})] = (A^\phi)_{kl}. \qquad \text{(A.7)}\end{aligned}$$

Therefore,

$$A^\alpha[\gamma^i|\gamma^j] = \begin{cases} A^\phi & i = j \\ 0 & i \neq j. \end{cases} \qquad \text{(A.8)}$$

□

Lemma A.2. *Let* $P = [D_0^T, D_1^T, \ldots, D_{L-1}^T]$, $E_\beta = I[\beta|\cdot]$, *and* E_ψ *be defined as in Eq. (5.47). Then,*

$$P^T E_\beta^T E_\beta P = E_\psi^T E_\psi. \tag{A.9}$$

Proof. The two sides of Eq. (A.9) can be written in block matrix form as

$$P^T E_\beta^T E_\beta P = \begin{bmatrix} D_0 E_\beta^T E_\beta D_0^T & D_0 E_\beta^T E_\beta D_1^T & \cdots & D_0 E_\beta^T E_\beta D_{L-1}^T \\ D_1 E_\beta^T E_\beta D_0^T & D_1 E_\beta^T E_\beta D_1^T & \cdots & D_1 E_\beta^T E_\beta D_{L-1}^T \\ \vdots & \vdots & \ddots & \vdots \\ D_{L-1} E_\beta^T E_\beta D_0^T & D_{L-1} E_\beta^T E_\beta D_1^T & \cdots & D_{L-1} E_\beta^T E_\beta D_{L-1}^T \end{bmatrix}, \tag{A.10}$$

and

$$E_\psi^T E_\psi = \begin{bmatrix} E_{\psi 0}^T E_{\psi 0} & 0 & \cdots & 0 \\ 0 & E_{\psi 1}^T E_{\psi 1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & E_{\psi L-1}^T E_{\psi L-1} \end{bmatrix}, \tag{A.11}$$

where every block is $C \times C$. The elements of $E_{\psi^i}^T E_{\psi^i}$ are

$$(E_{\psi^i}^T E_{\psi^i})_{kl} = \begin{cases} H^{\psi^i}[\mathcal{S}^{-1}(k, \mathbf{C})] & k = l \\ 0 & k \neq l, \end{cases} \tag{A.12}$$

and the *(i,j)*th block of $P^T E_\beta^T E_\beta P$ can be written as

$$D_i E_\beta^T E_\beta D_j^T = \mathrm{diag}(\mathbf{H}^\beta)[\gamma^i|\gamma^j]. \tag{A.13}$$

The elements of $\mathrm{diag}(\mathbf{H}^\beta)$ are

$$(\mathrm{diag}(\mathbf{H}^\beta))_{kl} = \begin{cases} H^\beta[\mathcal{S}^{-1}(k, \mathbf{N})] & k = l \\ 0 & k \neq l, \end{cases} \tag{A.14}$$

and therefore the elements of $\mathrm{diag}(\mathbf{H}^\beta)[\gamma^i|\gamma^j]$ are

$$\begin{aligned} (\mathrm{diag}(\mathbf{H}^\beta)[\gamma^i|\gamma^j])_{kl} &= (\mathrm{diag}(\mathbf{H}^\beta))_{\gamma_k^i \gamma_l^k} \\ &= \begin{cases} H^\beta[\mathcal{S}^{-1}(\gamma_k^i, \mathbf{N})] & \gamma_k^i = \gamma_l^j \\ 0 & \gamma_k^i \neq \gamma_l^j. \end{cases} \end{aligned} \tag{A.15}$$

If $i \neq j$, then $\gamma_k^i \neq \gamma_l^j$ and hence the *(k,l)*th element is zero. If $i = j$, then

$$\begin{aligned}(\mathrm{diag}(\mathbf{H}^\beta)[\gamma^i|\gamma^j])_{kl}\Big|_{i=j} &= \begin{cases} H^\beta[\mathcal{S}^{-1}(\gamma_k^i, \mathbf{N})] & \gamma_k^i = \gamma_l^i \\ 0 & \gamma_k^i \neq \gamma_l^i \end{cases} \\ &= \begin{cases} H^\beta[\mathcal{S}^{-1}(\mathcal{S}(\gamma_k^i, \mathbf{N}), \mathbf{N})] & k = l \\ 0 & k \neq l \end{cases} \\ &= \begin{cases} H^\beta[\gamma_k^i] & k = l \\ 0 & k \neq l, \end{cases}\end{aligned}$$

and by Eq. (A.5),

$$\begin{aligned} &= \begin{cases} H^\beta[\mathcal{S}^{-1}(i, \mathbf{L}) + \mathcal{S}^{-1}(k, \mathbf{C}) \circ \mathbf{L}] & k = l \\ 0 & k \neq l \end{cases} \\ &= \begin{cases} H^{\psi^i}[\mathcal{S}^{-1}(k, \mathbf{C})] & k = l \\ 0 & k \neq l \end{cases} \\ &= (E_{\psi^i}^T E_{\psi^i})_{kl}. \end{aligned} \tag{A.16}$$

Therefore,

$$D_i E_\beta^T E_\beta D_j^T = \begin{cases} E_{\psi^i}^T E_{\psi^i} & i = j \\ 0 & i \neq j. \end{cases} \tag{A.17}$$

□

Appendix B

Matlab code

A number of algorithms were developed in Chapters 5 and 6 of the thesis. Code for eleven MATLAB® functions demonstrating the implementation of these techniques is listed in this appendix. A mini-contents shown below includes section references and Fig. B.1 depicts the dependencies between functions.

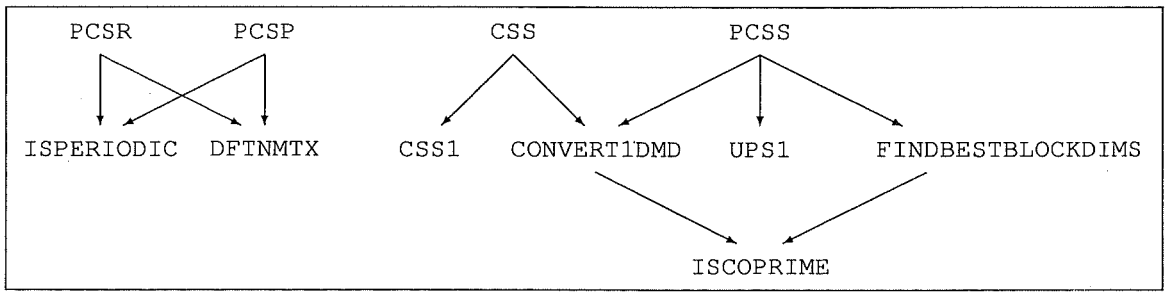


Figure B.1 Dependencies for the MATLAB® functions listed in Appendix B.

```
function [x,elim,tmetric,condnum] = pcsr(X,S,H,dB)
% PCSR Periodic Cartesian Sampling Reconstruction.
%
% [x,elim,tmetric,condnum] = PCSR(X,S,H,dB)
%
% Parameters:
% X - M-dimensional frequency domain data
% S - support mask for object (1=support,0=none)
% H - sampling mask for frequency domain data
%     (1=measured,0=not), DC=top left
% dB (optional) - SNR in dB present in the spectrum
%
% Returns:
% x - reconstructed spatial domain data
% elim - number of singular vectors eliminated
%        from each reconstruction
% tmetric - trace metric of each system of equations
% condnum - 2-norm condition number of each system of equations

% Note:
% - DFT is assumed to have 1/sqrt(prod(N)) on forward and inverse
%   (ie. different to MATLAB's FFT)

error(nargchk(3,4,nargin))
if ~isequal(size(X),size(S),size(H)),
    error('X, S and H must be the same size.')
end
N = size(X);
M = length(N);

%%% SETTING UP

% Find aliased sub-sequence
% (note the sqrt(prod(N)) factor to account for the
% different DFT definition in IFFTN)
x_tilde = ifftn(X.*H) * sqrt(prod(N));

% Find periodicity
[tf,C] = isperiodic(H);
L = N ./ C;

% Find sampling mask for the sub-sequences
idx = cell(1,M);
for I=1:M,
    idx{I} = [1:C(I)];
end
H_sub = H(idx{:});

% Create system matrix
% (note the prod(C) factor to account for the
% different DFT definition in DFTNMTX)
A = dftnmtx(C)' * diag(H_sub(:)) * dftnmtx(C) / prod(C);

% Precompute index sets to facilitate calculation in loop
[nn{1:M}] = ind2sub(C,1:prod(C));
for I=1:M,
    nn{I} = (nn{I}-1)*L(I) + 1;
end
J0 = sub2ind(N,nn{:});
[nn{1:M}] = ind2sub(L,1:prod(L));
J1 = sub2ind(N,nn{:});

%%% LOOP THROUGH AND SOLVE EACH SUB-SYSTEM

% Initialise arrays
x = zeros(N);
elim = zeros(L);
tmetric = zeros(L);
condnum = zeros(L);

% Main loop
h = waitbar(0,'Solving sub-systems...');
for I=1:prod(L),
    % Create index set of I'th sub-sequence
    J = J0 + J1(I) - 1;
    % Find sub-sequence support mask
    S_sub = S(J);
    % If support is not empty, solve system
    if sum(S_sub)>0,
        % Find aliased sub-sequence
        x_tilde_sub = x_tilde(J);
        % Find sub-sequence support set
        K = find(S_sub~=0);
        % Create system matrix (submatrix of A)
        B = A(:,K);
        % Compute SVD
        [U SS V] = svd(B,0);
        s = diag(SS);
        % Calculate conditioning information
        if s(length(s))>1e-15,
            condnum(I) = s(1)/s(length(s));
            tmetric(I) = sum(1./s.^2);
        else
            condnum(I) = inf;
            tmetric(I) = inf;
        end
        % Set elimination level for singular values
        if nargin<4,
            elim_level = 1e-10;
        else
            elim_level = max(s)/(10^(dB/10));
        end
        % Determine the number of singular values
        % to be eliminated
        r = sum(s > elim_level);
        elim(I) = length(s) - r;
        % Reconstruct sub-sequence
        y = zeros(prod(C),1);
        y(K) = V(:,1:r) * diag(ones(r,1)./s(1:r)) ...
            * U(:,1:r)' * x_tilde_sub(:);
        % Insert sub-sequence result into overall result
        x(J) = y(:);
    end
    waitbar(I/prod(L),h)
end
close(h)
```

```
function [U,sysrec,tmetric,condnum] = pcsp(S,H)
% PCSP Periodic Cartesian Sampling Prediction.
%
% [U,sysrec,tmetric,condnum] = PCSP(S,H)
%
% Parameters:
% S - support mask for object (1=support,0=none)
% H - sampling mask for frequency domain data
%     (1=measured,0=not), DC=top left
%
% Returns:
% U - image mask showing unrecoverable areas (1=unrecoverable,0=recoverable)
% sysrec - full recoverability of each system of equations (1=yes,0=no,-1=N/A)
% tmetric - trace metric of each system of equations
% condnum - 2-norm condition number of each system of equations

% Note:
% - DFT is assumed to have 1/sqrt(prod(N)) on forward and inverse
%   (ie. different to MATLAB's FFT)

error(nargchk(2,2,nargin))
if ~isequal(size(S),size(H)),
    error('S and H must be the same size.')
end
N = size(S);
M = length(N);

%%% SETTING UP

% Find periodicity
[tf,C] = isperiodic(H);
L = N ./ C;

% Find sampling mask for the sub-sequences
idx = cell(1,M);
for I=1:M,
    idx{I} = [1:C(I)];
end
H_sub = H(idx{:});

% Create system matrix
% (note the prod(C) factor to account for the
% different DFT definition in DFTNMTX)
A = dftnmtx(C)' * diag(H_sub(:)) * dftnmtx(C) / prod(C);

% Precompute index sets to facilitate calculation in loop
[nn{1:M}] = ind2sub(C,1:prod(C));
for I=1:M,
    nn{I} = (nn{I}-1)*L(I) + 1;
end
J0 = sub2ind(N,nn{:});
[nn{1:M}] = ind2sub(L,1:prod(L));
J1 = sub2ind(N,nn{:});

%%% LOOP THROUGH AND ANALYSE EACH SUB-SYSTEM

% Initialise arrays
U = zeros(N);
sysrec = -ones(L);
tmetric = zeros(L);
condnum = zeros(L);

% Main loop
h = waitbar(0,'Analysing sub-systems...');
for I=1:prod(L),
    % Create index set of I'th sub-sequence
    J = J0 + J1(I) - 1;
    % Find sub-sequence support mask
    S_sub = S(J);
    % If support is not empty, analyse system
    if sum(S_sub)>0,
        % Find sub-sequence support set
        K = find(S_sub~=0);
        % Create system matrix (submatrix of A)
        B = A(:,K);
        % Compute SVD
        s = svd(B);
        % Determine whether fully recoverable
        sysrec(I) = s(length(s))>1e-10;
        % Calculate conditioning information
        if s(length(s))>1e-15,
            condnum(I) = s(1)/s(length(s));
            tmetric(I) = sum(1./s.^2);
        else
            condnum(I) = inf;
            tmetric(I) = inf;
        end
        % Compute nullspace of reduced row echelon form
        % (use of RREF is slower but less chance of numerical error)
        Z = null(rref(B,1e-10),'r');
        % Check recoverability of elements in subsequence
        % (loop through all nullspace vectors; if the same element in all vectors is
        % zero then that element is recoverable)
        y = zeros(prod(C),1);
        for Q=1:size(Z,2),
            y(K) = y(K) | abs(Z(:,Q))>1e-10;
        end
        U(J) = y(:);
    else
        sysrec(I) = -1;
    end
    waitbar(I/prod(L),h)
end
close(h)
```

```
function H = css(S,p)
% CSS Cartesian sample selection.
%
% H = CSS(S,p)
%
% Parameters:
% S - support mask for object (1=support,0=none)
% p - number of samples to place
%
% Returns:
% H - sampling mask for frequency domain data
%     (1=measured,0=not), DC=top left
```

```
error(nargchk(2,2,nargin))
N = size(S);
M = length(N);

ONE_D = 0;
if M==2,
    if N(1)==1 | N(2)==1,
        ONE_D = 1;
    end
end

if ONE_D,
    N = max(N);
    beta = find(S);
    alpha = css1(N,beta,p);
    H = zeros(size(S));
    H(alpha+1) = 1;
else
    S1 = convert1dmd(N,S,1);
    alpha = css1(prod(N),find(S1),p);
    H1 = zeros(size(S1));
    H1(alpha+1) = 1;
    H = convert1dmd(N,H1);
end
```

```
function alpha = css1(N,beta,p)
% CSS1 Cartesian sample selection 1-D.
%
% alpha = CSS1(N,beta,p)
%
% Parameters:
% N - order of 1-D problem
% beta - support set (row vector, subset of [0..N-1])
% p (optional) - number of sampling locations to select
%                (default=number of samples in beta)
%
% Returns:
% alpha - sampling set (row vector, length p subset of [0..N-1])

error(nargchk(2,3,nargin))
if prod(size(N))~=1 | size(beta,1)~=1,
    error('N must be a scalar and beta must be a row vector.')
end
q = length(beta);
if nargin<3,
    p = q;
end

%%% SETTING UP

% Check whether to use the complementary size to speed up selection
USE_COMPLEMENT = (p > N/2);
if USE_COMPLEMENT,
    beta = setdiff([0:N-1],beta);
    beta = beta - beta(1);              % ensure first element is zero
    p = N - p;
    q = N - q;
end

% Initialise sampling set and possible choices
alpha = [0];
poss = [1:N-1];
Nalpha = 1;
Nposss = N - 1;

%% Note on calculation of metrics:

% With the other sample locations already selected, inside the loop only the
% incremental metric need be computed.  The term due to beta may be
% precomputed and placed in the LUT.  To avoid use of ABS() in the loop body,
% the LUT has values for both negative and positive differences.

%% Precompute LUT for factor metric

% Create list of pairs of samples to compare
beta_pairs = nchoosek(1:q,2);
% Find all distances between sample locations in beta
delta_beta = abs(beta(beta_pairs(:,2)) - beta(beta_pairs(:,1)));
% Determine the minimum distance (ie. d or N-d)
min_delta_beta = min(delta_beta,N - delta_beta);
% Create table of how many of distances of a particular size
Ndelta = floor(N/2);
DELTA_beta = histc(min_delta_beta,[1:Ndelta]);
% Find factors of N (excluding 1 and N)
x = find(mod(N,2:floor(sqrt(N)))==0) + 1;
f = unique([x fliplr(N./x)]);
Nf = length(f);
% Create term in formula due to beta
term_beta = zeros(1,Nf);
for I=1:Nf,
    term_beta(I) = sum(DELTA_beta(N/f(I)*[1:floor(f(I)/2)]));
end
% Create LUT
LUT_FM = zeros(Ndelta,1);
for I=1:Ndelta,
    for J=1:Nf,
        if floor(I/f(J))==I/f(J),
            LUT_FM(I) = LUT_FM(I) + term_beta(J);
        end
    end
end
LUT_FM(N-1:-1:Ndelta+1,:) = LUT_FM(1:ceil(N/2)-1,:);
LUT_FM = [flipud(LUT_FM);0;LUT_FM];

%% Precompute LUT for trace squared metric

% Create LUT
LUT_TSM = zeros(Ndelta,1);
if q > 1,
    for I=1:Ndelta,
        LUT_TSM(I) = sum(cos(2*pi/N*I*[1:Ndelta]).*DELTA_beta);
    end
end
LUT_TSM(N-1:-1:Ndelta+1) = LUT_TSM(1:ceil(N/2)-1);
LUT_TSM = [flipud(LUT_TSM);0;LUT_TSM];
```

```
%%% LOOP THROUGH AND SELECT (p-1) ELEMENTS OF ALPHA

for I=1:p-1,
    % Find all new differences between sample locations that would result
    % from each possible sample location
    delta = repmat(poss',1,Nalpha) - repmat(alpha,Nposs,1);

    %% Use factor metric

    % Compute metric using LUT
    metric = sum(LUT_FM(delta + N),2);
    % Find possible locations that result in the minimum factor metric
    FMset = find(abs(metric-min(metric))<1e-5);

    %% Use trace squared metric

    % Compute metric using LUT
    % (only need consider subset of new differences determined by factor metric)
    metric = sum(reshape(LUT_TSM(delta(FMset,:) + N),length(FMset),Nalpha),2);
    % Find possible locations that result in the minimum trace squared metric
    TSMset = find(abs(metric-min(metric))<1e-5);

    %% Make final random/arbitrary choice

    K = 1;

    %% Compute updates

    J = FMset(TSMset(K));              % find index of new sample location
    alpha = [alpha poss(J)];           % add new location to set
    poss = poss([1:J-1 J+1:Nposs]);    % remove location from possibilities
    Nalpha = Nalpha + 1;
    Nposs = Nposs - 1;
end

%%% FINAL TOUCHES

% Order the sampling set
alpha = sort(alpha);

% Restore the complement if necessary
if USE_COMPLEMENT,
    alpha = setdiff([0:N-1],alpha);
    alpha = alpha - alpha(1);          % ensure first element is zero
end
```

```
function [H,C,p,possdims] = pcss(S,lim,oversamp)
% PCSS Periodic Cartesian sample selection.
%
% [H,C,p,possdims] = PCSS(S,lim,oversamp)
%
% Parameters:
% S - support mask for object (1=support,0=none)
% lim - 2-element vector giving limits of total number of elements in repeated block
% oversamp (optional) - number of samples to oversample by in each block (default=0)
%
% Returns:
% H - sampling mask for frequency domain data
%     (1=measured,0=not), DC=top left
% C - dimensions of block size
% p - number of samples per block
% possdims - matrix of possible dimensions: columns are dimensions,
%            required number of samples per block, reduction in sampling density

error(nargchk(2,3,nargin))
N = size(S);
M = length(N);

%%% FIND BEST BLOCK SIZE

[C p possdims] = findbestblockdims(S,lim);

% FIND UNIVERSAL PATTERN

if nargin>2,
    p = p + oversamp;
    alpha = ups1(prod(C),p);
else
    alpha = ups1(prod(C),p);
end

%%% CREATE SAMPLING MASK

H1D = zeros(1,prod(C));
H1D(alpha+1) = 1;
Hblock = convert1dmd(C,H1D);
H = repmat(Hblock,N./C);
```

```
function alpha = ups1(N,p)
% UPS1 Universal pattern selection 1-D.
%
% alpha = UPS1(N,p)
%
% Parameters:
% N - order of 1-D problem
% p - number of sampling locations to select
%
% Returns:
% alpha - universal pattern sampling set
%         (row vector, length p subset of [0..N-1])

error(nargchk(2,2,nargin))
if prod(size(N))~=1 | prod(size(p))~=1,
    error('N and p must be scalars.')
end

%%% SETTING UP

% Check whether to use the complementary size to speed up selection
USE_COMPLEMENT = (p > N/2);
if USE_COMPLEMENT,
    p = N - p;
```

```
end

% Initialise sampling set and possible choices
alpha = [0];
poss = [1:N-1];
Nalpha = 1;
Nposs = N - 1;

%% Note on calculation of metrics:

% With the other sample locations already selected, inside the loop only the
% incremental factor metric need be computed.  However, the nature of the
% repeated distances metric means that a table of distances must be stored
% and the maximum found each time.  To avoid use of ABS() in the loop body,
% the factor metric LUT has values for both negative and positive differences.

%% Precompute LUT for factor metric

% Find factors of N (excluding 1 and N)
x = find(mod(N,2:floor(sqrt(N)))==0) + 1;
f = unique([x fliplr(N./x)]);
Nf = length(f);
% Create LUT
LUT_FM = zeros(N-1,1);
for I=1:Nf,
    LUT_FM(f(I):f(I):N-1) = LUT_FM(f(I):f(I):N-1) + 1;
end
LUT_FM = [flipud(LUT_FM);0;LUT_FM];

%% Initialise table for repeated distances metric updates

table_RDM = zeros(1,N-1);

%%% LOOP THROUGH AND SELECT (p-1) ELEMENTS OF ALPHA

for I=1:p-1,
    % Find all new differences between sample locations that would result
    % from each possible sample location
    delta = repmat(poss',1,Nalpha) - repmat(alpha,Nposs,1);

    %% Use factor metric

    % Compute metric using LUT
    metric = sum(LUT_FM(delta + N),2);
    % Find possible locations that result in the minimum factor metric
    FMset = find(abs(metric-min(metric))<1e-5);

    %% Use repeated distances metric

    % Only need consider subset of new differences determined by factor metric
    delta2 = delta(FMset,:);
    % Wrap negative distances onto 1..N-1
    delta2(find(delta2<1)) = delta2(find(delta2<1)) + N;
    % Create index set for use in RD table
    J = delta2 + repmat([0:length(FMset)-1]',1,Nalpha)*(N-1);
    % Create RD table using previously stored table and the new distances
    RD = repmat(table_RDM(:),1,length(FMset));
    RD(J) = RD(J) + 1;
    % Compute metric using RD table
    metric = max(RD(1:ceil(N/2)-1,:) + RD(N-1:-1:floor(N/2)+1,:),[],1);
    % Find possible locations that result in the minimum repeated distances metric
    RDMset = find(abs(metric-min(metric))<1e-5);

    %% Make final random/arbitrary choice

    K = 1;

    %% Compute updates

    table_RDM = RD(:,RDMset(K))';       % update RDM table using new sample location
    J = FMset(RDMset(K));               % find index of new sample location
    alpha = [alpha poss(J)];            % add new location to set
    poss = poss([1:J-1,J+1:Nposs]);     % remove location from possibilities
    Nalpha = Nalpha + 1;
    Nposs = Nposs - 1;
end

%%% FINAL TOUCHES

% Order the sampling set
alpha = sort(alpha);

% Restore the complement if necessary
if USE_COMPLEMENT,
    alpha = setdiff([0:N-1],alpha);
    alpha = alpha - alpha(1);           % ensure first element is zero
end
```

```
function Hout = convert1dmd(N,Hin,flag)
% CONVERT1DMD Convert mask between 1-D and M-D.
%
% Hout = CONVERT1DMD(N,Hin,flag)
%
% Parameters:
% N - row vector of dimensions of the M-D system
% Hin - 1-D or M-D sampling mask
% flag (optional) - indicates whether to use extra factor in conversion (default=0)
%
% Returns:
% Hout - M-D or 1-D sampling mask

error(nargchk(2,3,nargin))

Nin = size(Hin);
Min = length(Nin);
if Min==2,
    if Nin(1)==1 | Nin(2)==1,
        Min = 1;
    end
end
if nargin<3,
    flag = 0;
end

if Min==1,
    N = N(:);
    M = length(N);
    if prod(N)~=length(Hin),
```

```
        error('1-D to M-D: new dimensions must allow as many elements as in Hin.')
    end
    if ~iscoprime(N),
        error('1-D to M-D: dimensions must be pairwise relatively prime.')
    end
    % Find 1-D sample set
    alpha = find(Hin) - 1;
    % Use additional factor if required
    if flag,
        % Find conversion factor
        factor = sum(prod(N(nchoosek(1:M,M-1)),2));
        % Find multiplicative inverse of the factor
        [g c d] = gcd(prod(N),factor);
        ifactor = mod(d,prod(N));
        % Multiply sample set by inverse of the factor
        alpha = alpha * ifactor;
    end
    % Compute each dimension of new sample set
    alphaj = cell(1,M);
    for I=1:M,
        alphaj{I} = mod(alpha,N(I));
    end
    % Convert to subscripts starting from 1
    for I=1:M,
        alphaj{I} = alphaj{I} + 1;
    end
    % Convert subscripts into an array index
    J = sub2ind(N.',alphaj{:});
    % Create M-D sampling mask
    Hout = zeros(N.');
    Hout(J) = 1;
else
    N = size(Hin);
    N = N(:);
    M = length(N);
    if ~isequal(N.',Nin),
        error('M-D to 1-D: dimensions must be consistent with size of Hin.')
    end
    if ~iscoprime(N),
        error('1-D to M-D: dimensions must be pairwise relatively prime.')
    end
    % Find subscripts
    [alphaj{1:M}] = ind2sub(N.',find(Hin));
    % Convert to subscripts starting from 0
    for I=1:M,
        alphaj{I} = alphaj{I} - 1;
    end
    % Solve for alpha using Chinese Remainder Theorem
    NN = prod(N)./N;
    b = zeros(1,M);
    for I=1:M,
        [g c d] = gcd(NN(I),N(I));
        b(I) = mod(c,N(I));
    end
    alpha = zeros(length(alphaj{1}),1);
    for I=1:M,
        alpha = alpha + NN(I)*b(I)*alphaj{I};
    end
    alpha = mod(alpha,prod(N));
    % Use additional factor if required
    if flag,
        % Find conversion factor
        factor = sum(prod(N(nchoosek(1:M,M-1)),2));
        % Multiply sample set by factor
        alpha = mod(alpha*factor,prod(N));
    end
    % Create 1-D sampling mask
    Hout = zeros(1,prod(N));
    Hout(alpha+1) = 1;
end
```

```
function [C,p0min,possdims] = findbestblockdims(S,lim)
% FINDBESTBLKSIZE Find the block dimensions that give the biggest reduction in sampling
% density for the given support, constrained by the min/max total samples per block.
%
% [C,p0min,possdims] = FINDBESTBLOCKDIMS(S,lim)
%
% Parameters:
% S - support mask for object (1=support,0=none)
% lim - 2-element vector: [min max] number of elements per block
%
% Returns:
% C - dimensions of block size
% p0min - minimum number of samples required per block
% possdims - matrix of possible dimensions: columns are dimensions,
%            required number of samples per block, reduction in sampling density

error(nargchk(2,2,nargin))
if prod(size(lim))~=2,
    error('lim must be a 2-element vector.')
end

%%% SETTING UP

N = size(S);
M = length(N);

MAX_ENTRIES = 1000;                         % maximum possibilities for possdims matrix

best_reduction = 0;                         % initialise best reduction
best_C = N;                                 % initialise best dimensions
best_p0min = 0;                             % initialise best p0min
possdims = zeros(MAX_ENTRIES,M+2);          % matrix of possible block dimensions
num_entries = 0;                            % number of entries in possdims matrix

%%% FIND BLOCK DIMENSIONS TO TEST

%% Find all possible block dimensions

% Find factors of N(1)
x = find(mod(N(1),1:floor(sqrt(N(1))))==0);
f = unique([x fliplr(N(1)./x)]);

% Loop through remaining dimensions to create all combinations of block dimensions
Cposs = f;
for I=2:M,
    % Find factors of N(I)
    x = find(mod(N(I),1:floor(sqrt(N(I))))==0);
```

```
    f = unique([x fliplr(N(I)./x)]);
    % Update possible block dimenions
    Cposs = [kron(ones(1,length(f)),Cposs);kron(f,ones(1,size(Cposs,2)))];
end
Cposs = Cposs';

%% Reduce possible block dimensions to those that are within range

Cposs = Cposs(find(prod(Cposs,2)>=lim(1) & prod(Cposs,2)<=lim(2)),:);

%% Reduce possible block dimensions to those that are relatively prime

coprime_mask = zeros(size(Cposs,1),1);
for I=1:size(Cposs,1),
    coprime_mask(I) = iscoprime(Cposs(I,:));
end
Cposs = Cposs(find(coprime_mask),:);

%%% TEST POSSIBLE BLOCK DIMENSIONS

for I=1:size(Cposs,1),
    C = Cposs(I,:);

    %% Find p0min for this block dimensions

    % Find periodicity factor
    L = N ./ C;
    % Create index set to reorder support into subsupports
    [nn{1:M}] = ind2sub(C,1:prod(C));
    for I=1:M,
        nn{I} = (nn{I}-1)*L(I) + 1;
    end
    J0 = sub2ind(N,nn{:});
    [nn{1:M}] = ind2sub(L,1:prod(L));
    J1 = sub2ind(N,nn{:});
    J = repmat(J0,1,prod(L)) + kron(J1,ones(1,prod(C))) - 1;
    % Determine number of samples in each subsupport
    qi = sum(reshape(S(J),prod(C),prod(L)),1);
    % Compute minimum p0
    p0min = max(qi);

    %% Compare the result with previous block dimensions

    this_reduction = (prod(C) - p0min) / prod(C);
    num_entries = num_entries + 1;
    possdims(num_entries,:) = [C p0min this_reduction];
    % Update current best if better reduction
    if this_reduction > best_reduction,
        best_reduction = this_reduction;
        best_C = C;
        best_p0min = p0min;
    % If equal reduction, check whether smaller number of elements per block
    elseif this_reduction == best_reduction,
        if prod(C) < prod(best_C),
            best_reduction = this_reduction;
            best_C = C;
            best_p0min = p0min;
        end
    end
end
```

```
%%% SET RESULTS

C = best_C;
p0min = best_p0min;
possdims = possdims(1:num_entries,:);
```

```
function [TF,C] = isperiodic(X)
% ISPERIODIC True for periodic arrays.
%
% [TF,C] = ISPERIODIC(X)
%
% Parameters:
% X - M-dimensional array to determine periodicity of
%
% Returns:
% TF - boolean result (1=true,0=false)
% C - vector of periodicities in each dimension
%     (same size as size(X))

error(nargchk(1,1,nargin))
N = size(X);
M = length(N);

% Initialise in case no periodicity found
C = N;

% Check for periodicity in each dimension
for K=1:M,
    % Create strings to use in EVAL()
    sprefix = 'X(';
    for J=1:K-1,
        sprefix = strcat(sprefix,':,');
    end
    ssuffix = '';
    if K<M,
        ssuffix = strcat(ssuffix,',');
    end
    for J=K+1:M,
        ssuffix = strcat(ssuffix,':');
        if J<M,
            ssuffix = strcat(ssuffix,',');
        end
    end
    ssuffix = strcat(ssuffix,')');
    s = strcat('abs(',sprefix,'(I-1)*c+1:I*c',ssuffix,'-',...
        sprefix,'1:c',ssuffix,')');
    % Look for periodicity with period c
    for c=1:ceil(N(K)/2),
        % N must be a multiple of c
        if rem(N(K),c)==0,
            same = 1;
            % Check each N/c length for repetition
            for I=2:N(K)/c,
                chk = eval(s);
                % If no repetition found, stop checking
                if sum(chk(:))~=0,
                    same = 0;
```

```
                break
            end
        end
        % If periodicity was found, save results and stop
        % looking (smallest possible period will be found
        % as c is tested increasingly)
        if same==1,
            C(K) = c;
            break
        end
    end
  end
end
TF = ~isequal(C,N);
```

```
function W = dftnmtx(N)
% DFTNMTX M-D Fourier matrix.
%
% W = DFTNMTX(N)
%
% Parameters:
% N - vector of dimensions
%
% Returns:
% W - the prod(N) x prod(N) M-D Fourier matrix

error(nargchk(1,1,nargin))
N = N(:);
M = length(N);

W = 1;
for I=1:M,
    W = kron(dftmtx(N(I)),W);
end
error(nargchk(1,1,nargin))
N = N(:);
M = length(N);

W = 1;
for I=1:M,
    W = kron(dftmtx(N(I)),W);
end
error(nargchk(1,1,nargin))
N = N(:);
```

```
M = length(N);

W = 1;
for I=1:M,
    W = kron(dftmtx(N(I)),W);
end
```

```
function TF = iscoprime(X)
% ISCOPRIME True if two or more numbers are co-prime.
%
% TF = ISCOPRIME(X)
%
% Parameters:
% X - two or more positive integers to determine whether co-prime
%
% Returns:
% TF - Boolean result
%
% Definition:
% - two or more numbers are coprime if they have no factors in common other than 1

error(nargchk(1,1,nargin))
if prod(size(X))<2,
   error('At least two numbers must be considered.')
end
if sum(X<1)~=0 | sum(floor(X)~=X)~=0,
   error('Numbers must be positive integers.')
end

n = prod(size(X));
% Two are more numbers are coprime if they are pairwise coprime, so each pair of
% numbers needs to be checked

% Create list of pairs of numbers to compare
pairs = nchoosek(1:n,2);

% Loop through pairs of numbers
TF = 1;
for I=1:size(pairs,1),
   if gcd(X(pairs(I,1)),X(pairs(I,2)))~=1,
      TF = 0;
      break;
   end
end
```

References

[AB02] Nitin Aggarwal and Yoram Bresler. "Minimum redundancy non-uniform sampling in MRI using statistical criteria," in *Proceedings of the International Society of Magnetic Resonance in Medicine (ISMRM)*, Honolulu, Hawai'i, USA, May 2002, p. 2363.

[AIGBD94] Ender Ayanoğlu, Chih-Lin I, R. D. Gitlin, and Israel Bar-David. "Analog diversity coding to provide transparent self-healing communication networks," *IEEE Transactions on Communications*, vol. 42, no. 1, pp. 110–118, January 1994.

[APSW89] M. O. Ahmad, E. I. Plotkin, M. N. S. Swamy, and J. D. Wang. "An iterative procedure for uniform resampling of a 2-D signal from its nonequally-spaced samples," in *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS)*, Portland, Oregon, USA, May 1989, vol. 2, pp. 1087–1090.

[Bar90] Stephen Barnett. *Matrices: Methods and Applications*. Clarendon Press, Oxford, UK, 1990.

[BB98] Mario Bertero and Patrizia Boccacci. *Introduction to inverse problems in imaging*. Institute of Physics Publications, Bristol, UK, 1998.

[Ben92] John J. Benedetto. "Irregular sampling and frames," in *Wavelets—A Tutorial in Theory and Applications*, C. K. Chui, Ed., pp. 445–507. Academic Press, Boston, Massachusetts, USA, 1992.

[BF96] Yoram Bresler and Ping Feng. "Spectrum-blind minimum rate sampling and reconstruction of 2-D multiband signals," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, Lausanne, Switzerland, September 1996, vol. 1, pp. 701–704.

[BIG74] Adi Ben-Israel and Thomas N. E. Greville. *Generalized Inverses: Theory and Applications*. John Wiley & Sons, New York, USA, 1974.

[BPW02] John J. Benedetto, Alexander M. Powell, and Hui-Chuan Wu. "MRI signal reconstruction by Fourier frames on interleaving spirals," in *Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI)*, Washington, DC, USA, July 2002, pp. 717–720.

[Bra65] Ronald Newbold Bracewell. *The Fourier transform and its applications.* McGraw-Hill, New York, USA, 1965.

[Bri88] E. Oran Brigham. *The fast Fourier transform and its applications.* Prentice Hall, Englewood Cliffs, New Jersey, USA, 1988.

[Bro75] W. N. Brouw. "Aperture synthesis," in *Methods in Computational Physics*, B. Alder, S. Fernbach, and M. Rotenberg, Eds., vol. 14, pp. 131–175. Academic Press, New York, USA, 1975.

[BSAG99] Peter Börnert, Hermann Schomberg, Bernd Aldefeld, and Jan Groen. "Improvements in spiral MR imaging," *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 9, pp. 29–41, 1999.

[BW99] John J. Benedetto and Hui-Chuan Wu. "A multidimensional irregular sampling algorithm and applications," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Phoenix, Arizona, USA, March 1999, vol. 4, pp. 2039–2042.

[CA87] David Shi Chen and Jan P. Allebach. "Analysis of error in reconstruction of two-dimensional signals from irregularly spaced samples," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-35, no. 2, pp. 173–180, February 1987.

[Cal93] Paul T. Callaghan. *Principles of Nuclear Magnetic Resonance Microscopy.* Clarendon Press, Oxford, UK, 1993.

[Car58] H. Y. Carr. "Steady-state free precession in nuclear magnetic resonance," *Physical Review*, vol. 112, pp. 1693–1701, 1958.

[Cha88] Tony F. Chan. "An optimal circulant preconditioner for Toeplitz systems," *SIAM Journal on Scientific and Statistical Computing*, vol. 9, no. 4, pp. 766–771, July 1988.

[CLW95] T. J. Connolly, K. A. Landman, and L. R. White. "On Gerchberg's method for the Fourier inverse problem," *Journal of Australian Mathematical Society, Series B*, vol. 37, pp. 26–44, 1995.

[CM90] Kwan F. Cheung and Robert J. Marks II. "Imaging sampling below the Nyquist density without aliasing," *Journal of the Optical Society of America A*, vol. 7, no. 1, pp. 92–105, January 1990.

[CM98] Hyeokho Choi and David C. Munson, Jr. "Analysis and design of minimax-optimal interpolators," *IEEE Transactions on Signal Processing*, vol. 46, no. 6, pp. 1571–1579, June 1998.

[CM00] Hyeokho Choi and David C. Munson, Jr. "Stochastic formulation of bandlimited signal interpolation," *IEEE Transactions on Circuits and Systems—II: Analog and Digital Signal Processing*, vol. 47, no. 1, pp. 82–85, January 2000.

[CO94] Tony F. Chan and Julia A. Olkin. "Circulant preconditioners for Toeplitz-block matrices," *Numerical Algorithms*, vol. 6, pp. 89–101, 1994.

[Coh78] Daniel I. A. Cohen. *Basic Techniques of Combinatorial Theory*. John Wiley & Sons, New York, USA, 1978.

[CS92] Donald W. Chakeres and Petra Schmalbrock. *Fundamentals of Magnetic Resonance Imaging*. Williams & Wilkins, Baltimore, Maryland, USA, 1992.

[CT65] James W. Cooley and John W. Tukey. "An algorithm for the machine calculation of complex Fourier series," *Mathematics of Computation*, vol. 19, pp. 297–301, April 1965.

[Dav79] Philip J. Davis. *Circulant Matrices*. John Wiley & Sons, New York, USA, 1979.

[dB64] N. G. de Bruijn. "Pólya's theory of counting," in *Applied Combinatorial Mathematics*, E. F. Beckenbach, Ed., chapter 5, pp. 144–184. John Wiley & Sons, New York, USA, 1964.

[DCA+02] B. Desplanques, J. Cornelis, E. Achten, R. Van de Walle, and I. Lemahieu. "Iterative reconstruction of magnetic resonance images from arbitrarly samples in k-space," *IEEE Transactions on Nuclear Science*, vol. 49, no. 5, pp. 2268–2273, October 2002.

[DS52] R. J. Duffin and A. C. Schaeffer. "A class of nonharmonic Fourier series," *Transactions of the American Mathematical Society*, vol. 72, pp. 341–366, 1952.

[DWD01] Brian Dale, M. Wendt, and J. L. Duerk. "A rapid look-up table method for reconstructing MR images from arbitrary K-space trajectories," *IEEE Transactions on Medical Imaging*, vol. 20, no. 3, pp. 207–217, March 2001.

[FB96] Ping Feng and Yoram Bresler. "Spectrum-blind minimum-rate sampling and reconstruction of multiband signals," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Atlanta, USA, May 1996, pp. 1688–1691.

[Fer94] Paulo Jorge S. G. Ferreira. "Interpolation and the discrete Papoulis-Gerchberg algorithm," *IEEE Transactions on Signal Processing*, vol. 42, no. 10, pp. 2596–2606, October 1994.

[FG92] Hans G. Feichtinger and Karlheinz Gröchenig. "Iterative reconstruction of multivariate band-limited functions from irregular sampling values," *SIAM Journal on Mathematical Analysis*, vol. 23, no. 1, pp. 244–261, January 1992.

[FHM86] J. Frahm, A. Haase, and D. Matthaei. "Rapid NMR imaging of dynamic processes using the FLASH technique," *Magnetic Resonance in Medicine*, vol. 3, pp. 321–327, 1986.

[FS01] Jeffrey A. Fessler and Bradley P. Sutton. "A min-max approach to the multidimensional nonuniform FFT: application to tomographic image reconstruction," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, Thessaloniki, Greece, October 2001, vol. 1, pp. 706–709.

[Fud89] M. Fuderer. "Ringing artefact reduction by an efficient likelihood improvement method," in *SPIE Proceedings*, Paris, France, April 1989, vol. 1137, pp. 84–90.

[GDMC+96] D. Graveron-Demilly, G. J. Marseille, Y. Crémillieux, S. Cavassila, and D. van Ormondt. "SRS-FT, a Fourier imaging method based on sparse radial scanning and bayesian estimation," *Journal of Magnetic Resonance, Series B*, vol. 112, pp. 119–123, 1996.

[Ger74] R. W. Gerchberg. "Super-resolution through error energy reduction," *Optica Acta*, vol. 21, no. 9, pp. 709–720, 1974.

[GN93] G. H. Glover and D. C. Noll. "Consistent projection reconstruction (CPR) techniques for MRI," *Magnetic Resonance in Medicine*, vol. 29, pp. 345–351, 1993.

[GR00] Yun Gao and Stanley J. Reeves. "Optimal k-space sampling in MRSI for images with a limited region of support," *IEEE Transactions on Medical Imaging*, vol. 19, no. 12, pp. 1168–1178, December 2000.

[GR01] Yun Gao and Stanley J. Reeves. "Fast *k*-space sample selection in MRSI with a limited region of support," *IEEE Transactions on Medical Imaging*, vol. 20, no. 9, pp. 868–876, September 2001.

[Gro92] Karlheinz Gröchenig. "Reconstruction algorithms in irregular sampling," *Mathematics of Computation*, vol. 59, no. 199, pp. 181–194, July 1992.

[GV96] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. John Hopkins University Press, Baltimore, Maryland, USA, 3rd edition, 1996.

[HJ85] Roger A. Horn and Charles R. Johnson. *Matrix Analysis.* Cambridge University Press, Cambridge, UK, 1985.

[HKP97] Richard D. Hoge, Remi K. S. Kwan, and G. Bruce Pike. "Density compensation functions for spiral MRI," *Magnetic Resonance in Medicine*, vol. 38, pp. 117–128, 1997.

[Hou72] G. N. Hounsfield. "A method of and apparatus for examination of a body by radiation such as X-ray or gamma radiation," *British Patent No. 1283915*, 1972.

[Hu94] Xiaoping Hu. "On the "keyhole" technique," *Journal of Magnetic Resonance Imaging*, vol. 4, no. 2, pp. 231, 1994.

[Hun95] B. R. Hunt. "Super-resolution of images: algorithms, principles, performance," *International Journal of Imaging Systems and Technology*, vol. 6, pp. 297–304, 1995.

[HW56] G. H. Hardy and E. M. Wright. *An Introduction to the Theory of Numbers.* Clarendon Press, Oxford, UK, 3rd edition, 1956.

[HW99] Cormac Herley and Ping Wah Wong. "Minimum rate sampling and reconstruction of signals with arbitrary frequency support," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1555–1564, July 1999.

[IJ93] Emmanuel C. Ifeachor and Barrie W. Jervis. *Digital Signal Processing: A Practical Approach.* Addison-Wesley, Harlow, UK, 1993.

[Jen88] Yih-Chyun Jenq. "Digital spectra of nonuniformly sampled signals: Fundamentals and high-speed waveform digitizers," *IEEE Transactions on Instrumentation and Measurement*, vol. 37, no. 2, pp. 245–251, June 1988.

[Jer77] Abdul J. Jerri. "The Shannon sampling theorem—its various extensions and applications: A tutorial review," *Proceedings of the IEEE*, vol. 65, no. 11, pp. 1565–1597, November 1977.

[JHM+93] R. A. Jones, O. Haraldseth, T. B. Müller, P. A. Rinck, and A. N. Øksendal. "K-space substitution: A novel dynamic imaging technique," *Magnetic Resonance in Medicine*, vol. 29, pp. 830–834, 1993.

[JMNM91] John I. Jackson, Craig H. Meyer, Dwight G. Nishimura, and Albert Macovski. "Selection of a convolution function for Fourier inversion using gridding," *IEEE Transactions on Medical Imaging*, vol. 10, no. 3, pp. 473–478, September 1991.

[Jon86] M. C. Jones. "The discrete Gerchberg algorithm," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-34, no. 3, pp. 624–626, June 1986.

[JR81] Anil K. Jain and Surendra Ranganath. "Extrapolation algorithms for discrete signals with application in spectral estimation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-29, no. 4, pp. 830–845, August 1981.

[KFC95] Kevin F. King, Thomas K. F. Foo, and Carl R. Crawford. "Optimizaed gradient waveforms for spiral scanning," *Magnetic Resonance in Medicine*, vol. 34, pp. 156–160, 1995.

[Lan67] H. J. Landau. "Necessary density conditions for sampling and interpolation of certain entire functions," *Acta Mathematica*, vol. 117, pp. 37–52, February 1967.

[Lau73] Paul C. Lauterbur. "Image formation by induced local interations: examples employing nuclear magnetic resonance," *Nature*, vol. 242, pp. 190–191, 1973.

[Lia00] Jan-Ray Liao. "Real-time image reconstruction for spiral MRI using fixed-point calculation," *IEEE Transactions on Medical Imaging*, vol. 19, no. 7, pp. 690–698, July 2000.

[LL00] Zhi-Pei Liang and Paul C. Lauterbur. *Principles of Magnetic Resonance Imaging: A Signal Processing Perspective*. IEEE Press, New York, USA, 2000.

[Lue84] David G. Luenberger. *Linear and Nonlinear Programming*. Addison-Wesley, Reading, Massachusetts, USA, 2nd edition, 1984.

[Mar86] Robert J. Marks II. "Multidimensional-signal sample dependency at Nyquist densities," *Journal of the Optical Society of America A*, vol. 3, no. 2, pp. 268–273, February 1986.

[MdBF+96] G. J. Marseille, R. de Beer, M. Fuderer, A. F. Mehlkopf, and D. van Ormondt. "Nonuniform phase-encode distributions for MRI scan time reduction," *Journal of Magnetic Resonance, Series B*, vol. 111, pp. 70–75, 1996.

[MFdB+94] G. J. Marseille, M. Fuderer, R. de Beer, A. F. Mehlkopf, and D. van Ormondt. "Reduction of MRI scan time through nonuniform sampling and edge-distribution modeling," *Journal of Magnetic Resonance, Series B*, vol. 103, pp. 292–295, 1994.

[MGP99] B. Madore, G. H. Glover, and N. J. Pelc. "Unaliasing by Fourier-encoding the overlaps using the temporal dimension (UNFOLD), applied to cardiac imaging and fMRI," *Magnetic Resonance in Medicine*, vol. 42, pp. 813–828, 1999.

[Mil70] K. Miller. "Least-squares method for ill-posed problems with a prescribed bound," *SIAM Journal on Mathematical Analysis*, vol. 1, pp. 52–74, February 1970.

[Mil95] R. P. Millane. "Aspects of image reconstruction from nonuniform samples," in *SPIE Proceedings*, San Diego, California, USA, July 1995, vol. 2570, pp. 385–391.

[Nas76] M. Z. Nashed. "On moment-discretization and least-squares solutions of linear integral equations of the first kind," *Journal of Mathematical Analysis and Applications*, vol. 53, pp. 359–366, 1976.

[Nis95] Dwight G. Nishimura. *Principles of Magnetic Resonance Imaging*. Standford University, add-sanfrancisco-us, 1995. (lecture notes).

[NL99] S. K. Nagle and D. N. Levin. "Multiple region MRI," *Magnetic Resonance in Medicine*, vol. 41, pp. 774–786, 1999.

[NL01] S. K. Nagle and D. N. Levin. "Creating a library of generalized Fourier sampling patterns for irregular 2D regions of support," *Magnetic Resonance in Medicine*, vol. 46, pp. 624–627, 2001.

[Nol97] Douglas C. Noll. "Multishot rosette trajectories for spectrally selective MR imaging," *IEEE Transactions on Medical Imaging*, vol. 16, no. 4, pp. 372–377, August 1997.

[Now99] Robert D. Nowak. "Wavelet-based Rician noise removal for magnetic resonance imaging," *IEEE Transactions on Image Processing*, vol. 8, no. 10, pp. 1408–1419, October 1999.

[O'S85] J. O'Sullivan. "A fast sinc function gridding algorithm for Fourier inversion in computer tomography," *IEEE Transactions on Medical Imaging*, vol. MI-4, no. 4, pp. 200–207, December 1985.

[Pap75] Athanasios Papoulis. "A new algorithm in spectral analysis and band-limited extrapolation," *IEEE Transactions on Circuits and Systems*, vol. CAS-22, no. 9, pp. 735–742, September 1975.

[PK83] S. X. Pan and Avinash C. Kak. "A computational study of reconstruction algorithms for diffraction tomography: Interpolation versus filtered backpropagation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-31, no. 5, pp. 1262–1275, October 1983.

[PMBdM84] E. R. Pike, J. G. McWhirter, M. Bertero, and C. de Mol. "Generalised information theory for inverse problems in signal processing," *IEE Proceedings, Part F*, vol. 131, no. 6, pp. 660–667, October 1984.

[PWSB99] Klaas P. Pruessmann, Markus Weiger, Markys B. Scheidegger, and Peter Boesiger. "SENSE: Sensitivity encoding for fast MRI," *Magnetic Resonance in Medicine*, vol. 42, pp. 952–962, 1999.

[Red96] Don Redmond. *Number theory: an introduction*. Marcel Dekker, Inc., New York, USA, 1996.

[RH95] Stanley J. Reeves and Larry P. Heck. "Selection of observations in signal reconstruction," *IEEE Transactions on Signal Processing*, vol. 43, no. 3, pp. 788–791, March 1995.

[Rus87] C. K. Rushforth. "Signal restoration, functional analysis, and Fredholm integral equations of the first kind," in *Image Recovery: Theory and Application*, Henry Stark, Ed., chapter 1, pp. 1–27. Academic Press, Orlando, Florida, USA, 1987.

[SA87] Ken D. Sauer and Jan P. Allebach. "Iterative reconstruction of band-limited images from nonuniformly spaced samples," *IEEE Transactions on Circuits and Systems*, vol. CAS-34, no. 12, pp. 1497–1506, December 1987.

[San94] Irwin W. Sandberg. "The reconstruction of band-limited signals from nonuniformly spaced samples," *IEEE Transactions on Circuits and Systems—I: Fundamental Theory and Applications*, vol. 41, no. 1, pp. 64–66, January 1994.

[Sar01] Gordon E. Sarty. "Direct reconstruction of non-cartesian k-space data using a nonuniform fast Fourier transform," *Magnetic Resonance in Medicine*, vol. 45, pp. 908–915, 2001.

[SDWW99] Kyle A. Salem, Jeffrey L. Duerk, Michael Wendt, and David L. Wilson. "Evaluation of keyhole MR imaging with a human visual response model," in *SPIE Proceedings*, San Diego, California, USA, February 1999, vol. 3663, pp. 232–242.

[Sez90] M. Ibrahim Sezan. "A peak detection algorithm and its application to histogram-based image data reduction," *Computer Vision, Graphics, and Image Processing*, vol. 49, no. 1, pp. 36–51, January 1990.

[Sha49] Claude E. Shannon. "Communication in the presence of noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, January 1949.

[She94a] Ram G. Shenoy. "Nonuniform sampling of signals and applications," in *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS)*, London, UK, May 1994, vol. 2, pp. 181–184.

[She94b] Jonathan Richard Shewchuk. "An introduction to the conjugate gradient method without the agonizing pain," (Paper available by anonymous FTP to `WARP.CS.CMU.EDU` in directory `quake-papers/` as the file `painless-conjugate-gradient.ps`, or through the URL `http://www.cs.cmu.edu/~quake-papers/painless-conjugate-gradient.ps`), 1994.

[SL74] L. A. Shepp and B. F. Logan. "The Fourier reconstruction of a head section," *IEEE Transactions on Nuclear Science*, vol. NS-21, no. 3, pp. 21–43, June 1974.

[SM97] Daniel K. Sodickson and Warren J. Manning. "Simultaneous acquisition of spatial harmonics (SMASH): Fast imaging with radiofrequency coil arrays," *Magnetic Resonance in Medicine*, vol. 38, pp. 591–603, 1997.

[SN00] Hossein Sedarat and Dwight G. Nishimura. "On the optimality of the gridding reconstruction algorithm," *IEEE Transactions on Medical Imaging*, vol. 19, no. 4, pp. 306–317, April 2000.

[SP61] D. Slepian and H. O. Pollak. "Prolate spheroidal wave functions, Fourier analysis and uncertainty—I," *Bell System Technical Journal*, vol. 40, no. 1, pp. 43–64, January 1961.

[SS87] M. Ibrahim Sezan and Henry Stark. "Applications of convex projection theory to image recovery in tomography and related areas," in *Image Recovery: Theory and Application*, Henry Stark, Ed., chapter 11, pp. 415–462. Academic Press, Orlando, Florida, USA, 1987.

[Str86] Gilbert Strang. "A proposal for Toeplitz matrix calculations," *Studies in Applied Mathematics*, vol. 74, pp. 171–176, 1986.

[Str88] Gilbert Strang. *Linear Algebra and its Applications*. Harcourt Brace Jovanovich College Publishers, Orlando, Florida, USA, 3rd edition, 1988.

[Str97] Thomas Strohmer. "Computationally attractive reconstruction of bandlimited images from irregular samples," *IEEE Transactions on Image Processing*, vol. 6, no. 4, pp. 540–548, April 1997.

[Tar97] Andrzej Tarczynski. "Sensitivity of signal reconstruction," *IEEE Signal Processing Letters*, vol. 4, no. 7, pp. 192–194, July 1997.

[TB74] A. R. Thompson and R. N. Bracewell. "Interpolation and Fourier transformation of fringe visibilities," *Astronomical Journal*, vol. 79, no. 1, pp. 11–24, January 1974.

[Tho72] R. C. Thompson. "Principal submatrices IX: Interlacing inequalities for singular values of submatrices," *Linear Algebra and its Applications*, vol. 5, pp. 1–12, 1972.

[Tik63] A. N. Tikhonov. "Solution of incorrectly formulated problems and the regularization method," *Soviet Mathematics Doklady*, vol. 4, pp. 1035–1038, 1963.

[Uns00] Michael Unser. "Sampling—50 years after Shannon," *Proceedings of the IEEE*, vol. 88, no. 4, pp. 569–587, April 2000.

[VB98] Raman Venkataramani and Yoram Bresler. "Further results on spectrum blind sampling of 2D signals," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, Chicago, USA, October 1998, vol. 2, pp. 752–756.

[VB00] Raman Venkataramani and Yoram Bresler. "Perfect reconstruction formulas and bounds on aliasing error in sub-Nyquist nonuniform sampling of multiband signals," *IEEE Transactions on Information Theory*, vol. 46, no. 6, pp. 2173–2183, September 2000.

[VB01] Raman Venkataramani and Yoram Bresler. "Optimal sub-Nyquist nonuniform sampling and reconstruction for multiband signals," *IEEE Transactions on Signal Processing*, vol. 49, no. 10, pp. 2301–2313, October 2001.

[VBM+00] Rik Van de Walle, Harrison H. Barrett, Kyle J. Myers, Maria I. Altbach, Bart Desplanques, Arthur F. Gmitro, Jan Cornelis, and Ignace Lemahieu. "Reconstruction of MR images from data acquired on a general nonregular grid by pseudoinverse calculation," *IEEE Transactions on Medical Imaging*, vol. 19, no. 12, pp. 1160–1167, December 2000.

[VL90] P. P. Vaidyanathan and Vincent C. Liu. "Efficient reconstruction of band-limited sequences from nonuniformly decimated versions by use of polyphase filter banks," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 11, pp. 1927–1936, November 1990.

[vVBD+93] Joop J. van Vaals, Marijn E. Brummer, W. Thomas Dixon, Hans H. Tuithof, Hans Engels, Rendon C. Nelson, Brigid M. Gerety, Judith L. Chezmar, and Jacques A. den Boer. ""keyhole" method for accelerating imaging of contrast agent uptake," *Journal of Magnetic Resonance Imaging*, vol. 3, no. 4, pp. 671–675, 1993.

[WCL+02] F. T. A. W. Wajer, A. Coron, R. Lethmate, J. A. C. van Osch, L. T. Martinez, D. Graveron-Demilly, and D. van Ormondt. "Accelerated Bayesian MR image reconstruction," in *Proceedings of the International Society of Magnetic Resonance in Medicine (ISMRM)*, Honolulu, Hawai'i, USA, May 2002, p. 2422.

[WdBF+98] F. T. A. W. Wajer, R. de Beer, M. Fuderer, A. F. Mehlkopf, and D. van Ormondt. "Bayesian image reconstruction from an arbitrarily and sparsely sampled K-space without gridding and density correction," in *Proceedings of the International Society of Magnetic Resonance in Medicine (ISMRM)*, Sydney, Australia, April 1998, p. 667.

[Wel91] Emo Welzl. "Smallest enclosing disks (balls and ellipsoids)," in *New Results and New Trends in Computer Science*, H. Maurer, Ed., vol. 555 of *Lecture Notes in Computer Science*, pp. 359–370. Springer-Verlag, Berlin, Germany, 1991.

[Win92] Duncan J. Wingham. "The reconstruction of a band-limited function and its Fourier transform from a finite number of samples at arbitrary locations by singular value decomposition," *IEEE Transactions on Signal Processing*, vol. 40, no. 3, pp. 559–570, March 1992.

[WN94] David O. Walsh and Pamela A. Nielsen-Delaney. "Direct method for superresolution," *Journal of the Optical Society of America A*, vol. 11, no. 2, pp. 572–579, February 1994.

[WSB+01] F. T. A. W. Wajer, G. H. L. A. Stijnman, M. Bourgeois, D. Graveron-Demilly, and D. van Ormondt. "Magnetic resonance image reconstruction from nonuniformly sampled k-space data," in *Nonuniform Sampling: Theory and Practice*, Farokh Marvasti, Ed., chapter 10, pp. 439–478. Kluwer Academic, New York, USA, 2001.

[Xia95] Changbai Xiao. "Reconstruction of bandlimited signal with lost samples at its Nyquist rate—the solution to a nonuniform sampling problem," *IEEE Transactions on Signal Processing*, vol. 43, no. 4, pp. 1008–1009, April 1995.

[Yen56] J. L. Yen. "On nonuniform sampling of bandwidth-limited signals," *IRE Transactions on Circuit Theory*, vol. 3, pp. 251–257, 1956.

[You78] D. C. Youla. "Generalized image restoration by the method of alternating orthogonal projections," *IEEE Transactions on Circuits and Systems*, vol. CAS-25, no. 9, pp. 694–702, September 1978.

[YS90] Shu-jen Yeh and Henry Stark. "Iterative and one-step reconstruction from nonuniform samples by convex projections," *Journal of the Optical Society of America A*, vol. 7, no. 3, pp. 491–499, March 1990.

[YW82] D. C. Youla and H. Webb. "Image restoration by the method of convex projections: Part 1-theory," *IEEE Transactions on Medical Imaging*, vol. MI-1, no. 2, pp. 81–94, October 1982.